University of Alberta

Cloning and characterization of the peptide synthetase gene cluster involved in the nonribosomal biosynthesis of fusaricidin-type antifungal antibiotics in *Paenibacillus polymyxa* PKB1

by

Jingru Li ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**

in

Microbiology and Biotechnology

Department of Biological Sciences

Edmonton, Alberta
Fall 2007

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33013-5
Our file Notre référence
ISBN: 978-0-494-33013-5

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Paenibacillus polymyxa (formerly *Bacillus polymyxa*) PKB1 produces fusaricidin-type antifungal antibiotics that inhibit the growth of *Leptosphaeria maculans*, a plant pathogenic fungus causing blackleg disease of canola. Fusaricidin consists of a guanidino-modified β-hydroxy fatty acid linked to a cyclic hexapeptide with four residues present in the D-configuration.

Using a reverse genetic approach based on the conserved motifs of nonribosomal peptide synthetases, the entire fusaricidin biosynthetic gene cluster (*fus*) was cloned and sequenced, and spans 32.4 kb including an open reading frame (*fusA*) encoding a six-module peptide synthetase. The second, fourth and fifth modules of fusaricidin synthetase (FusA) each contain an epimerization domain, consistent with the incorporation of D-amino acids in these positions of fusaricidin. However, the sixth module, corresponding to D-Ala, lacks an epimerization domain. The sixth adenylation domain of FusA was produced at high levels in *Escherichia coli* and shown to activate D-Ala specifically, providing evidence for direct selection and activation of a D-amino acid by a typical prokaryotic peptide synthetase. The *fus* cluster also includes genes presumably involved in biosynthesis, modification, and activation of the lipid moiety of fusaricidin. However, no genes for regulation, resistance, or transport functions were encountered.

To confirm the involvement of *fusA* in fusaricidin production, a modified PCR-targeting mutagenesis protocol was developed to create a *fusA* mutation on the chromosome of PKB1. A DNA fragment internal to *fusA* was replaced by a gene disruption cassette containing two antibiotic resistance markers for independent selection of apramycin resistance in *E. coli* and chloramphenicol resistance in *P. polymyxa*. The inclusion of an *oriT* site in the disruption cassette allowed efficient transfer of the inactivated *fusA* allele into *P. polymyxa* by intergeneric conjugation from *E. coli*. Targeted disruption of *fusA* led to the complete loss of the antifungal activity against *L. maculans*, suggesting that *fusA* plays an essential role in the nonribosomal synthesis of fusaricidin. The boundaries of the *fus* gene cluster were determined using the same mutagenesis strategy.

The results presented in this thesis provide the basis for genetic manipulation of fusaricidin production in *P. polymyxa* PKB1, and furthermore, for construction of novel antibiotics by combinatorial biosynthesis.

Acknowledgements

I would like to thank Dr. Susan Jensen for providing me with the opportunity to work on this project, and for all her kindness, advice and guidance throughout my graduate study. I thank Alberta Research Council for their continuing interest in this project, and in particular, Dr. Saleh Shah for providing the degenerate oligonucleotide primer B1 and sequences of PCR-amplified peptide synthetase gene fragments from *P. polymyxa* PKB1. In addition, I would like to thank Dr. Sui-Lam Wong for his critical comments on my thesis.

A special thank to Annie Wong for patiently answering my questions, and for her technical assistance with sequencing part of the fosmid inserts. I thank Nathan Zelyas for his friendship and helpful discussions with me about gene expression and protein purification. I would also like to thank Dr. Mike Harrington for sharing his broad knowledge in many areas of biology.

I thank Huirong Chen, Dongfang Hui, and Wei Wang for becoming my best friends, and for their endless encouragement and support in my hard times. In addition, I would like to express my sincere gratitude to my parents for their unconditional love of me and confidence in my abilities to succeed, without which this work would not have been completed.

Finally, I would like to acknowledge the financial support for my research and graduate study by the Natural Sciences and Engineering Research Council of Canada, and the Department of Biological Sciences at the University of Alberta.

Table of Contents

Page

1. Introduction.. 1

1.1 Canola.. 1

1.2 *Leptosphaeria maculans*.. 2

1.3 *Paenibacillus polymyxa*.. 4

1.4 Nonribosomal peptide synthetases.. 6

  1.4.1 Catalytic domains of modular peptide synthetases.................... 8

    1.4.1.1 Adenylation domain: the primary determinant of substrate specificity.. 10

    1.4.1.2 Thiolation domain: the peptidyl carrier protein.................... 11

    1.4.1.3 Condensation domain: the nonribosomal elongation factor........... 12

    1.4.1.4 Thioesterase domain: the nonribosomal hydrolase or cyclase......... 13

    1.4.1.5 Modifying domains: tailoring enzymes embedded in NRPSs......... 15

  1.4.2 Rational design of NRPSs towards construction of novel peptides............ 17

1.5 Nonribosomally synthesized lipopeptides.. 20

1.6 Hybrid peptide-polyketide synthetase systems....................................... 21

1.7 Main objectives of this thesis... 24

2. Materials and methods.. 43

2.1 Strains of bacteria and fungi, plasmids, reagents and supplies...................... 43

2.2 *E. coli* procedures.. 44

  2.2.1 Growth conditions.. 44

  2.2.2 Preparation of *E. coli* electrocompetent cells and electroporation............ 45

2.2.3 Preparation of colony lifts... 46

2.3 *Paenibacillus polymyxa* procedures... 47

2.3.1 Growth conditions and media.. 47

2.3.2 Purification of genomic DNA from *P. polymyxa*............................. 48

2.3.3 Plasmid isolation from *P. polymyxa*... 48

2.3.4 Preparation of *P. polymyxa* electrocompetent cells and electroporation...... 50

2.3.5 Transfer of DNA into *P. polymyxa* by intergeneric conjugation.............. 51

2.3.6 Antifungal activity bioassay.. 51

2.3.7 HPLC analysis of fusaricidin production... 52

2.4 DNA manipulations.. 53

2.4.1 Gel purification of DNA fragments.. 53

2.4.2 Polymerase chain reaction (PCR).. 54

2.4.3 Transfer of DNA fragments onto nylon membranes........................... 56

2.4.4 Labeling of DNA probes.. 57

2.4.5 Hybridization of Southern and colony blots...................................... 57

2.4.6 DNA sequencing and analysis.. 58

2.4.7 Preparation of *P. polymyxa* genomic libraries.................................. 59

2.4.8 Creation of an Apra^{R}Cm^R*oriT* disruption cassette.................... 60

2.4.9 PCR-targeted gene disruption... 61

2.5 Protein analysis... 62

2.5.1 Overproduction and purification of the internal adenylation domain FusA-A6... 62

2.5.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)....................... 64

2.5.3 ATP-PPi exchange assay... 64

3. Results and discussion.. **74**

3.1 Analysis of the peptide synthetase gene fragments amplified from

P. polymyxa PKB1 75
3.2 Screening of the first SuperCos-1 genomic library of *P. polymyxa* PKB1 76
3.2.1 The first screening of PKB1 genomic library, with the B1J6-10 fragment 77
3.2.2 The second screening of PKB1 genomic library, with the B1J6-17 fragment 79
3.2.3 Sequence analysis of the insert DNA of cosmid Col-8 81
3.2.4 Sequence analysis of cosmid Col-19 83
3.2.4.1 Identification of a putative fusaricidin synthetase gene, *orf7* 83
3.2.4.2 Cloning of a continuous chromosomal fragment downstream of *orf7* 84
3.3 Preparation and screening of a second SuperCos-1 genomic library of *P. polymyxa* PKB1 86
3.4 Preparation and screening of a fosmid genomic library of *P. polymyxa* PKB1 87
3.5 Disruption of putative fusaricidin synthetase genes 89
3.5.1 Identification of selectable markers for use in *P. polymyxa* PKB1 90
3.5.2 Transformation of *P. polymyxa* PKB1 with an integrational vector carrying the *fus-asn* gene fragment 90
3.5.3 Insertional mutation of *fus-asn* via double crossover 92
3.5.4 Preparation of fusaricidin-nonproducer mutants using a modified Redirect© technology 94
3.5.4.1 Adaptation of the PCR-targeting protocol for *P. polymyxa* 94
3.5.4.2 Disruption of the putative peptide synthetase gene *orf7* on cosmid Col-19 97
3.5.4.3 Phenotype of the *orf7* mutants 99
3.6 Analysis of the fusaricidin biosynthetic gene cluster and the flanking regions 100
3.6.1 Modular organization of the fusaricidin synthetase 101

3.6.1.1 Condensation (C) domains.. 102

3.6.1.2 Epimerization (E) domains.. 104

3.6.1.3 Thiolation (T) domains.. 106

3.6.1.4 Adenylation (A) domains.. 106

3.6.2 Determination of the substrate specificity of the A domain in the sixth module of FusA.. 109

3.6.3 N-terminal lipidation of fusaricidin.. 110

3.6.4 The 3' boundary of the *fusA* gene cluster.. 111

3.6.5 The 5' boundary of the *fusA* gene cluster.. 111

3.6.5.1 Preparation of the *ymcC* mutant.. 112

3.6.5.2 Preparation of the *ktp* mutant.. 113

4. Conclusions, significance and future research........................ 169

4.1 Conclusions and significance.. 169

4.2 Proposed future research.. 172

5. References.. 175

6. Appendices.. 189

Appendix A: The nucleotide sequence of the insert DNA of cosmid Col-8............ 189

Appendix B: The nucleotide sequence of the insert DNA of cosmid Col-19............ 206

Appendix C: The nucleotide sequence of the fusaricidin biosynthetic gene cluster and the flanking regions from *P. polymyxa* PKB1........................ 221

List of Tables

Page

1.1 The conserved core motifs found in the common catalytic domains of nonribosomal peptide synthetases........ 42

2.1 Bacterial and fungal strains used in this study........ 66

2.2 Cloning vectors and recombinant plasmids used in this study........ 68

2.3 Oligonucleotide PCR primers used in this study........ 71

3.1 Predicted substrate specificities of the adenylation (A) domains derived from the PCR-amplified peptide synthetase gene fragments, based on the selectivity-conferring codes of NRPSs........ 161

3.2 Summary of the proposed functions of ORFs identified within the insert of cosmid Col-8........ 162

3.3 The substrate specificity-conferring codes of the adenylation domains derived from ORFs identified in the cosmid inserts of Col-8 and Col-19........ 163

3.4 Summary of the proposed functions of ORFs identified within the insert of cosmid Col-19........ 164

3.5 Summary of ORFs identified in the fusaricidin biosynthetic gene cluster and the flanking regions........ 166

3.6 The amino acid residues lining the substrate-binding pockets of the adenylation domains in fusaricidin synthetase FusA........ 168

List of Figures

Page

1.3.1 Primary structure of the fusaricidin-type lipopeptide antibiotics produced by *P. polymyxa*........ 26

1.4.1 Schematic representation of the module and domain organization of nonribosomal peptide synthetases........ 27

1.4.2 Schematic diagram of amino acid adenylation and subsequent thiolation during nonribosomal peptide synthesis........ 28

1.4.3 Extraction of the substrate specificity-conferring residues by alignment of the amino acid sequences in the region of core motifs A4 and A5........ 29

1.4.4 Posttranslational modification of the peptide synthetase (NRPS), from the inactive apo-form to the active holo-form, prior to peptide biosynthesis........ 30

1.4.5 Peptide bond formation between two amino acids that are thioester-bound to the thiolation (T) domains of two successive modules, mediated by the condensation (C) domain........ 31

1.4.6 A schematic diagram showing two termination strategies of nonribosomal peptide synthesis........ 32

1.4.7 Modifications of amino acid substrates being incorporated into nonribosomal peptides, catalyzed by optional modifying domains........ 33

1.4.8 Possible strategies to construct hybrid peptide synthetases for the synthesis of new peptide products........ 35

1.5.1 Nonribosomally synthesized cyclic lipopeptides produced by *Bacillus subtilis*... 37

1.6.1 Schematic diagram of catalytic functions of Type I PKS domains........ 39

1.6.2 Proposed modular organization of the hybrid NRPS-PKS synthetase encoded by the leinamycin biosynthetic gene cluster (*lnm*)........ 41

3.1.1 The PCR primers used to amplify fragments of the peptide synthetase genes from *P. polymyxa* PKB1 114

3.2.1 Southern analysis of cosmid Col-8 identified in the first screening of the SuperCos-1 genomic library of *P. polymyxa* PKB1 117

3.2.2 Southern analysis of the positive cosmids identified in the second screening of the first SuperCos-1 genomic library of *P. polymyxa* PKB1 118

3.2.3 ORF organization of the insert DNA from cosmid Col-8 119

3.2.4 The structures of commonly used polymyxin-type antibiotics produced by *P. polymyxa* 121

3.2.5 ORF organization of the insert DNA from cosmid Col-19 122

3.2.6 Schematic representation of the peptide synthetase gene comprising *orf7* from Col-19 and the PCR-amplified fragment TOPO-19 123

3.4.1 Southern analysis of the hybridizing fosmids identified in the pSMART-FOS genomic library of *P. polymyxa* PKB1 125

3.4.2 Southern analysis of the fosmids 4G9 and 6D11 identified in the pSMART-FOS genomic library of *P. polymyxa* PKB1 126

3.5.1 Southern analysis of the putative *fus-asn* single-crossover mutants 127

3.5.2 Gene disruption of the *fus-asn* fragment in *P. polymyxa* PKB1 via double crossover 128

3.5.3 Analysis of the putative *fus-asn* double-crossover mutants by colony PCR 130

3.5.4 Southern analysis of the putative *fus-asn* double-crossover mutants 131

3.5.5 Restriction analysis of the plasmid DNA isolated from chloramphenicol resistant (Cm^R) *P. polymyxa* exconjugants 132

3.5.6 Diagrammatic representation of the Apra^R*oriT* and Apra^{R}Cm^R*oriT* disruption cassettes 134

3.5.7 PCR amplification of the Apra^{R}Cm^R*oriT* gene disruption cassette for preparation of *orf7* mutants using a modified Redirect© technology.............. 135

3.5.8 Construction of the mutant cosmid Col-19 containing a disrupted copy of *orf7*...136

3.5.9 Gene disruption of the chromosomal copy of *orf7* in *P. polymyxa* PKB1......... 138

3.5.10 PCR verification of the *P. polymyxa orf7* mutants.................................. 140

3.5.11 Southern analysis of the *P. polymyxa orf7* mutants................................ 132

3.5.12 The antifungal activity of A4, A5, and A6 mutants carrying the *orf7* disruption compared to the activity of PKB1 strain................................ 144

3.5.13 HPLC analysis of methanol extracts from the wild-type and A4 mutant cultures.. 145

3.6.1 ORF arrangement of the fusaricidin biosynthetic gene cluster...................... 146

3.6.2 Module and domain organization of the fusaricidin synthetase encoded by *fusA* 147

3.6.3 Comparison of the conserved core motifs within the condensation (C) domains of FusA.. 148

3.6.4 Alignment of the amino acid sequences of six thiolation (T) domains in FusA... 149

3.6.5 Alignment of the amino acid sequences in core motifs A4 and A5 of FusA-A6 compared to other known A domains assigned to activate L- or D-Ala........... 150

3.6.6 Amino acid sequence of the adenylation (A) and thiolation (T) domains in the sixth module (FusA-M6) of fusaricidin synthetase FusA.......................... 151

3.6.7 Overeproduction and purification of His-tagged recombinant FusA-A6 protein 152

3.6.8 Relative activities of the purified FusA-A6 protein for various amino acid substrates determined using the ATP-PPi exchange assay.......................... 153

3.6.9 PCR-targeted mutagenesis of *ymcC* in *P. polymyxa* PKB1........................ 154

3.6.10 Verification of the *P. polymyxa ymcC* mutants by PCR reactions................ 156

3.6.11 Comparison of the antifungal activity of wild-type PKB1 with the *ymcC* and *ktp* mutants.......... 158

3.6.12 PCR analysis of the *P. polymyxa ktp* mutant.......... 159

List of Symbols

α	Alpha
β	Beta
δ	Delta
Δ	Deletion
ε	Epsilon
γ	Gamma
λ	Lambda bacteriophage
μ	Micro
Ω	Omega

List of Abbreviations

A	Adenylation domain / alanine
aa	Amino acid
aadA	Streptomycin resistance gene
ACP	Acyl carrier protein
ACV	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine
Ala	Alanine
amp	Ampicillin resistance gene
aac(3)IV	Apramycin resistance gene
Asn	Asparagine
AT	Acyl transferase domain
B	Asparagine or aspartic acid
BHI	Brain heart infusion medium
Blast	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
C	Condensation domain / cysteine
cat	Chloramphenicol resistance gene
Ci	Curie
Cy	Cyclization domain
D	Aspartic acid
DAB	2,4-diaminobutyric acid
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphat
dGTP	Deoxyguanosine triphosphat
DH	Dehydratase domain
dITP	Deoxyinosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
E	Epimerization domain / glutamic acid
EDTA	Ethylenediamine tetraacetic acid
ER	Enoylreduction domain

F	Phenylalanine / Farad
FMN	Flavin mononucleotide
FRT	FLP recognition target
g	Gram
G	Glycine
GB	Glucose broth
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
H	Histidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
I	Isoleucine
Ile	Isoleucine
IPTG	Isopropyl-β-D-thiogalactopyranoside
K	Lysine
kan	Kanamycin resistance gene
kb	Kilobase
kDa	KiloDalton
KR	Ketoreductase domain
KS	Ketosynthase domain
kV	KiloVolt
L	Leucine
LB	Luria-Bertani medium
Leu	Leucine
M	Methylation domain / molar / methionine
N	Asparagine
neo	Neomycin resistance gene
NRPS	Nonribosomal peptide synthetase
nt	Nucleotide
NT	Nick translation
OD	Optical density
ORF	Open reading frame

Orn	Ornithine
oriT	Origin of transfer
Ox	Oxidation domain
P	Proline
PAGE	Polyacrylamide gel electrophoresis
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDA	Potato dextrose agar
PDB	Potato dextrose broth
PEG	Polyethylene glycol
Phe	Phenylalanine
PKS	Polyketide synthase
4'-PP	4'-phosphopantetheine
Q	Glutamine
R	Reductase domain / arginine
RNase	Ribonuclease
rpm	Revolutions per minute
S	Serine
SAM	S-adenosylmethionine
SDS	Sodium dodecyl sulfate
Ser	Serine
SOB	Super optimal broth
SSC	Standard saline citrate
T	Thiolation domain / threonine
TCSS	Trypticase soy with starch medium
TE	Thioesterase domain / Tris-EDTA buffer
TEMED	N,N,N',N'-tetra-methylethylenediamine
Thr	Threonine
Tris	Tris(hydroxymethyl)-aminomethane
Trp	Tryptophan
Tyr	Tyrosine
U	Unit
UV	Ultraviolet

V	Valine / volt
Val	Valine
v/v	Volume per volume
W	Tryptophan
w/v	Weight per volume
x	Unknown amino acid
X-gal	5-bromo-4-chloro-3-β-D-galactopyranoside
Y	Tyrosine
Z	Glutamine or glutamic acid
ZOI	Zone of inhibition

1. Introduction

1.1 Canola

The history of canola as a crop goes back to oilseed rape or rapeseed, a member of the family Crucifereae and the genus *Brassica* along with 3,000 other species. Today, four species of *Brassica* are grown in many areas of the world for the production of animal feed and vegetable oil for human consumption: biennial or winter annual forms of *Brassica napus* in Europe and China, summer annual forms of *B. napus* or *B. rapa* in western Canada, *B. juncea* which predominates in India, and *B. carinata* in eastern Africa (Rimmer and Vandenberg, 1992). The close genetic relationship existing among *Brassica* species has enabled plant breeders to transfer useful characteristics from one species to another through cross hybridization. Canola (mainly *B. napus* and *B. rapa*) was initially bred in Canada through traditional plant breeding techniques in the 1970s. Compared with rapeseed plants, canola contains reduced levels of both glucosinolates, producing toxic byproducts upon hydrolysis, and erucic acid, a fatty acid that has been related to heart disease in humans when present at high concentrations. The word "canola" is derived from "Canadian oil, low acid", and it is now used worldwide to describe varieties of rapeseed, from which rapeseed oil with less than 2% erucic acid and rapeseed animal meal with less than 3 mg/g of glucosinolates are obtained (Vaisey-Genser and Eskin, 1982). Many insects, such as flea beetles, may infest canola at various growth stages. However, severe losses in canola yield are mainly associated with phoma stem canker (blackleg) disease caused by the fungus *Leptosphaeria maculans*, which has been documented for over 100 years. Blackleg is an endemic disease in the main rapeseed growing areas of Australia, Canada and Europe, but rare in the areas of oilseed rape grown in Scotland, India and China (Gugel and Petrie, 1992).

1.2 *Leptosphaeria maculans*

The soil-inhabiting fungus *Leptosphaeria maculans* (anamorph: *Phoma lingam*) causes blackleg, the most severe disease of *Brassica* species (particularly *B. napus* and *B. rapa*). Strains of *L. maculans* have been classified into two groups based on their cultural, biochemical and genetic characteristics: the highly virulent or aggressive group that grows slowly, causes damaging stem cankers, and produces phytotoxin sirodesmins (the causal agent of blackleg), and the weakly virulent or nonaggressive group that grows rapidly, causes less damage on plants, and fails to produce sirodesmins (Mcgee and Petrie, 1978; Williams and Fitt, 1999). Attempts to mate in vitro between these two groups of *L. maculans* have failed, and there has been no evidence of natural inter-strain mating, as well.

L. maculans can produce pycnidiospores asexually within pycnidia, and ascospores sexually within pseudothecia. Ascospores present on infested canola stubble act as the primary inoculum of blackleg disease. They are released during rainfall and spread by wind as aerosols, which usually coincides with the sowing period, resulting in infection of the emerging seedlings (Hall, 1992). Pycnidiospores are the secondary inocula, which are spread by rain splash onto the leaves of neighboring canola plants (Gugel and Petrie, 1992; Hall, 1992). Following initial infection of the leaves through stomata or wounds, the fungus grows intercellularly through the plant tissues as a biotroph, and spreads down the plant from the leaves to the stem and even to the roots. After this symptomless stage, the fungus induces plant cell death and degradation, and utilizes the degraded cellular components as nutrients to produce more pycnidia and pycnidiospores in the dead tissues, resulting in dark lesions on all infected areas. Stem infection can produce a girdling blackened canker along the stem, hence the so-called blackleg. This stem cankering causes premature ripening of the pods due to disrupted water and nutrient transport, and more severely lodging and death of the plants. Due to dramatic loss in oilseed production associated with blackleg, cultural, chemical and other methods have been used to control both the severity and spread of the disease.

Ascospores of *L. maculans* can survive in the infested land using the remaining rapeseed residues as nutrients. Therefore, a four-year crop rotation is recommended in most oilseed rape growing areas, which provides time (three years) for infested rapeseed residues to break down before the next successive oilseed rape crops (Gugel and Petrie, 1992). Degradation of infected residues can be facilitated by raking, burning or burying, and improved methods are being sought (Khangura and Barbetti, 2001). Sustainable management of phoma stem canker also requires breeding of rapeseed cultivars with resistance to *L. maculans* isolates at both seedling stage, to impede initial spreading from the leaves to the stem, and adult plant stage, to prevent the development of stem lesions. Blackleg resistant cultivars of *B. napus* have been developed by plant breeders in Europe, Australia and Canada (Gugel and Petrie, 1992; Rimmer and Vandenberg, 1992). In addition, extensive studies have been conducted to identify genetic sources of rapeseed resistance (Gugel and Petrie, 1992; Rimmer and Vandenberg, 1992). Genetic modification of oilseed rape through the introduction of resistance genes from other plants or genes coding for production of antifungal compounds, was also reported (Wang *et al.*, 1999). Furthermore, fungicidal seed and soil treatments, as well as foliar applications of fungicides, were adopted for control of *L. maculans* infection in different oilseed rape growing regions (Gugel and Petrie, 1992; Kharbanda, 1992). However, given environmental and health concerns about the use of chemicals, the development of natural and efficient biocontrol agents for blackleg disease is more desirable.

Kharbanda and Dahiya (1990) identified a strain of *Penicillium verrucosum* that inhibited the growth of *L. maculans* by producing a secondary metabolite called citrinin. However, the toxicity of citrinin on farm animals prevents its use as a biocontrol agent. More recently, an isolate of *Paenibacillus polymyxa* was obtained from canola stubble, producing antifungal peptides that are able to inhibit *L. maculans* in greenhouse trials (Beatty and Jensen, 2002). It was found that several fungicides and herbicides approved for use on canola crops in Canada had no effect on this *P. polymyxa* isolate, so it has the potential to be used as a direct biocontrol agent, either alone or in combination with conventional fungicides and herbicides.

1.3 *Paenibacillus polymyxa*

The free-living soil bacterium *Paenibacillus polymyxa* was a member of the genus *Bacillus*. It was regrouped into a new genus *Paenibacillus* (paeni- meaning almost) and described as the type species based on 16S rRNA sequence similarity (Ash *et al.*, 1993; Jurtshuk *et al.*, 1992). *P. polymyxa* is a Gram-positive, rod-shaped, motile organism, and the cells develop subterminal ellipsoidal endospores when nutrients become limited. As the species name suggests (poly-myxa meaning much-slime), *P. polymyxa* produces a large amount of extracellular levan in liquid or on solid media (Han, 1989; Shida *et al.*, 1997a, 1997b). When grown on carbohydrate-rich solid media, the colonies are usually coated with shiny, opaque capsular slime. This species appears to contain a circular genome of approximately 4.3 Mb with a G+C content of 45%. There has been only one report of *P. polymyxa* plasmid, in which a cryptic linear plasmid (pFS1) was isolated from *P. polymyxa* SCE2 strain (Rosado and Seldin, 1993).

P. polymyxa populations are often found in the rhizosphere of various crop plants (Heulin *et al.*, 1994; Holl *et al.*, 1988). They exhibit plant growth promoting properties such as nitrogen fixation, phosphate solubilization, and secretion of plant growth promoting compounds (Heulin *et al.*, 1994; Jisha and Alagawadi, 1996). Certain strains of *P. polymyxa* are able to enhance plant growth indirectly by stimulating symbiotic bacterial populations to grow in the rhizosphere (Petersen *et al.*, 1996). In addition, *P. polymyxa* displays antagonistic activity against plant pathogenic fungi by producing various degrading enzymes and antibiotic compounds (Dijksterhuis *et al.*, 1999; Heulin *et al.*, 1994).

P. polymyxa strains have been shown to produce a wide range of peptide antibiotics, which may provide them with a growth advantage in the competitive soil environment. The peptide metabolites that have been identified in various isolates of *P. polymyxa* are generally classified into two groups according to their antimicrobial activities. Members of the first group include the polymyxins (Katz and Demain, 1977), polypeptins (Sogn, 1976), jolipeptin (Ito and Koyama, 1972a, 1972b), gavaserin and saltavalin (Pichard *et al.*, 1995), which are typified by antibacterial activity against both Gram-negative and Gram-positive bacteria, as well as by the presence of the nonproteinogenic amino acid, 2,4-diaminobutyric acid (DAB). The second group consists of a single family of closely related peptides variously designated as gatavalin (Nakajima *et al.*, 1972), fusaricidins A to D (Kajimura and Kaneda, 1996, 1997), or LI-F antibiotics (twelve distinct compounds) (Kuroda *et al.*, 2000; Kurusu *et al.*, 1987), all of which contain an unusual fatty acid side chain, 15-guanidino-3-hydroxypentadecanoic acid. Their antagonist activity against fungi and Gram-positive bacteria with no effect on Gram-negative bacteria distinguishes this second group of peptides from the first. All peptides of this second group will hereafter be referred to as fusaricidins in this thesis. The general peptide sequence of the fusaricidins was determined as L-Thr-X_2-X_3-D-*allo*-Thr-X_5-D-Ala (Figure 1.3.1A). The β-hydroxy fatty acid is attached to the N-terminal L-Thr, via an amide linkage, and the peptide is cyclized by an ester bond between the carboxy group of the C-terminal D-Ala and the β-hydroxy group of the N-terminal L-Thr. The antimicrobial activity of fusaricidins has been shown to vary depending on the particular amino acids present at three variable positions in the peptide moiety.

In summary, *P. polymyxa*, a naturally soilborne bacterium, possesses several properties desirable in a biocontrol agent active against plant pathogenic fungi, including production of highly resistant endospores and various peptide antibiotics, as well as intrinsic resistance to several fungicides and herbicides approved for use on oilseed rape in Canada.

1.4 Nonribosomal peptide synthetases

Bioactive peptides that are produced by microorganisms represent a large group of diverse natural products, including antibiotics, immunosuppressive agents, toxins, enzyme inhibitors, and so on. Therefore, the investigation of their biosynthetic mechanisms is of great benefit to medicine, agriculture, and biological research. Some of these peptides, such as nisin and subtilin, are synthesized ribosomally through proteolytic processing and posttranslational modifications of the gene-encoded precursors. Yet another class of natural peptides, with remarkable structural diversity, was shown to be assembled by the nonribosomal peptide synthetases (NRPSs), the multi-enzyme complexes used as assembly lines directing the stepwise peptide elongation (Marahiel *et al.*, 1997; von Dohren *et al.*, 1997).

The nonribosomal mechanism of peptide biosynthesis was first reported as early as the 1970s by Lipmann and coworkers. The synthesis of some peptide antibiotics was first observed in cellular extracts that had been treated with ribonuclease, and subsequently it was noted that the cyclic antibiotics gramicidin S and tyrocidine from *Bacillus* appeared to be synthesized by large enzymes in a way similar to fatty acid synthesis (Lipmann, 1971; Lipmann *et al.*, 1971; Lipmann, 1973). In this nucleic acid-independent pathway, large enzyme complexes are used as both template and biosynthetic machinery to direct the incorporation of amino acids into peptide polymers. In contrast to ribosomal biosynthesis, nonribosomally produced peptides contain not only the common 20 amino acids but also nonproteinogenic residues, such as D-amino acids or carboxy acids. Additionally, a common feature of many nonribosomal peptides is their cyclic or branched cyclic structures that are modified by N-methylation, glycosylation, oxidative cross-linking, as well as the incorporation of fatty acids and small heterocyclic rings. These structural peculiarities clearly distinguish nonribosomally synthesized peptides from ribosomal peptides and proteins, and play an important role in proper interaction between peptide compounds and their specific molecular targets in the cell (Marahiel *et* *al.*, 1997; von Dohren *et al.*, 1997). Moreover, nonribosomal peptides are very often produced as homologous series instead of single compounds, suggesting the relatively broad specificity of this biosynthetic pathway compared to ribosomal peptide synthesis. Therefore, it was speculated that enzymatically-directed peptide synthesis may have existed prior to, but evolved in parallel to, the ribosomal mechanism (Lipmann, 1971).

1.4.1 Catalytic domains of modular peptide synthetases

By using both genetic and biochemical approaches, studies on the nonribosomal mechanism of peptide biosynthesis have advanced substantially in the past decades. Despite the structural diversity, synthesis of most nonribosomally assembled peptides employs a common "multiple carrier thiotemplate mechanism" (Lipmann, 1980). According to this model, amino acid substrates are first activated by the peptide synthetase enzymes, and subsequently transferred to the free thiol groups of multiple enzyme-bound cofactors of the 4'-phosphopantetheine (4'-PP) type. These 4'-PP cofactors act as flexible swinging arms that are able to transfer the thioester-bound amino acids from one site to the next on the peptide synthetases for optional modification or peptide bond formation. During the stepwise peptide elongation, all intermediates are covalently tethered to the protein template as thioesters via the 4'-PP cofactors.

It is now generally accepted that nonribosomal peptide synthetases (NRPSs) are organized into an assembly line of modules, each responsible for the incorporation of one amino acid substrate into the peptide product (Marahiel *et al.*, 1997; von Dohren *et al.*, 1997). Usually, the number and order of modules within the NRPSs directly corresponds to the number and sequence of amino acid residues in the synthesized peptide, since nonribosomal peptide synthesis generally proceeds in a colinear way from the N-terminus to the C-terminus. Peptide synthetases of this type are hence referred to as linear NRPSs. In contrast, peptides composed of dimers of two, or trimers of three, are synthesized by iterative NRPSs, whose modules and domains are used more than once during the peptide assembly. Nonlinear NRPSs represent a third biosynthetic strategy of nonribosomal peptide synthesis, in which the arrangement of modules and domains does not correspond to the peptide sequence (Mootz *et al.*, 2002).

Each module consists of structurally and enzymatically independent domains connected by flexible linker regions. Each domain fulfills a specific catalytic or carrier function during nonribosomal peptide synthesis, and domains retain enzymatic activities when expressed heterologously as isolated proteins. In most cases, the individual domains can be identified easily at the protein level by the presence of characteristic, consensus amino acid sequences (Table 1.1). It has also been shown that these core motifs contain highly conserved residues directly involved in the domains' functions. The adenylation domain, the thiolation domain and the condensation domain are three core domains that are necessary for the nonribosomal biosynthesis of the peptide backbone, and thus define a minimal elongation module (Figure 1.4.1) (Challis and Naismith, 2004). The initiation module or loading module usually lacks a condensation domain, while the most downstream termination module has an extra specialized C-terminal thioesterase domain for the release of the fully assembled peptide chain. In addition, modules may contain optional domains that introduce modifications of the amino acids being incorporated into the peptide product.

1.4.1.1 Adenylation domain: the primary determinant of substrate specificity

The first step in nonribosomal peptide synthesis is the selection and activation of a specific substrate by the adenylation (A) domain. The A domain, about 550 amino acids in length, catalyzes the activation of an amino acid as aminoacyl adenylate, resulting from transfer of AMP from ATP to the substrate in the presence of Mg^{2+} and release of pyrophosphate (Figure 1.4.2) (Dieckmann *et al.*, 1995). The corresponding reaction in ribosomal peptide synthesis is carried out by aminoacyl-tRNA synthetases, but these two enzyme families share no similarity in either sequence or structure (Eriani *et al.*, 1990; Rapaport *et al.*, 1985). In addition, compared with aminoacyl-tRNA synthetases, the NRPS A domains exhibit relaxed substrate specificity, capable of activating more than one amino acid.

The A domains belong to the superfamily of adenylate-forming enzymes that includes firefly luciferases and acyl-CoA ligases (Turgay *et al.*, 1992). The resolution of the crystal structures of two A domains, the Phe-activating A domain (PheA) of the gramicidin synthetase GrsA from *Bacillus brevis*, and the stand-alone A domain (DhbE) of the bacillibactin synthtase, activating aryl acid 2,3-dihydroxybenzoate (DHB), from *Bacillus subtilis*, provides detailed insight into the structural basis of substrate recognition and adenylation (Conti *et al.*, 1997; May *et al.*, 2002). These two enzymes share a similar folding pattern, composed of a large N-terminal subdomain and a small C-terminal subdomain that are linked by a hinge region. The active site, where the substrate binds, is located at the interface of the two subdomains. It was found that most of the conserved core motifs (A1-A10) reside surrounding the active site. This observation confirms the importance of these consensus motifs for the catalytic activity of the A domain, which had been demonstrated using site-directed mutagenesis and photoaffinity labeling experiments (Gocht and Marahiel, 1994; Pavela-Vrancic *et al.*, 1994). Comparison of the amino acid residues lining the substrate-binding pocket in PheA with the corresponding moieties in other A domains led to the identification of important residues involved in substrate recognition by A domains (Figure 1.4.3) (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999). Introduction of targeted mutations within the selectivity-conferring residues of PheA resulted in predicted alterations of the substrate specificity (Stachelhaus *et al.*, 1999). Therefore, these residues can be defined as the codons of nonribosomal peptide synthetases. Using the nonribosomal code, it is now possible to predict the substrate specificity of biochemically uncharacterized A domains simply by sequence analysis.

1.4.1.2 Thiolation domain: the peptidyl carrier protein

In the next step of nonribosomal biosynthesis, the aminoacyl adenylate intermediate is transferred to the free thiol group of the 4'-phosphopantetheine (4'-PP) cofactor, which is covalently bound to the thiolation (T) domain of about 80 residues, located directly downstream of the A domain (Figure 1.4.2). Since the T domain shows high similarity to the acyl carrier proteins (ACP) of polyketide synthases and fatty acid synthases, this domain is also referred to as peptidyl carrier protein (PCP) (Stachelhaus *et al.*, 1996; Weber *et al.*, 2000). The T domain is the only NRPS domain without its own autonomous catalytic function, but all aminoacyl or peptidyl intermediates tethered to the flexible 4'-PP cofactors can be transported to other catalytic domains for subsequent reactions. The 4'-PP cofactor is attached to the side chain of an invariant serine residue within the core sequence [LGG(HD)**S**(LI)] of the T domain. The posttranslational apo-to-holo modification of T domains by binding the 4'-PP cofactor, is catalyzed by dedicated 4'-PP transferases using coenzyme A (CoA) as a common substrate (Figure 1.4.4A), which occurs before the actual peptide biosynthesis (Lambalot *et al.*, 1996; Schlumbohm *et al.*, 1991). Genes encoding 4'-PP transferases in bacteria are often found adjacent to peptide synthetase genes.

1.4.1.3 Condensation domain: the nonribosomal elongation factor

Condensation (C) domains (about 450 amino acids in length) located N-terminal of A domains, are responsible for the stepwise N- to C- terminal elongation of the peptide chain. They catalyze peptide bond formation between the upstream T domain-bound amino acid or peptidyl chain and the downstream T domain-bound amino acid (Figure 1.4.5) (Stachelhaus *et al.*, 1998). Usually, the number of C domains within NRPSs is consistent with the number of peptide bonds in the synthesized peptide. Recently, by site-directed mutations in the tyrocidine synthetase TycB of ten residues highly conserved among eighty other C domains, a possible catalytic mechanism of the C domain was proposed, involving the Arg residue in the core motif C2 (**R**HExLRTxF), and the His and Asp residues in the core motif C3 (MH**H**xIS**D**G) (Bergendahl *et al.*, 2002).

The elucidation of the crystal structure of a freestanding C domain, VibH of the vibriobactin synthetase from *Vibrio cholerae*, revealed the presence of an active site at the interface of the two subdomains of VibH (Keating *et al.*, 2002). The cognate substrates of VibH, 2,3-dihydroxybenzoate and norspermidine, are proposed to enter the active site from opposite faces of the C domain, which correspond to the so-called donor site, accepting the amino acid or peptide chain tethered to the upstream T domain, and the acceptor site, accepting the amino acid tethered to the downstream T domain. Biochemical studies of various C domains indicated that the acceptor site exhibits selectivity towards both the stereochemistry (L- or D-configuration) and the side chain of cognate amino acids (Belshaw *et al.*, 1999; Ehmann *et al.*, 2000). Although the donor site appears to be more tolerant of substrates with noncognate side chains, stereoselectivity towards the upstream thioester-bound amino acid, or the C-terminal amino acid of the growing peptidyl chain, was also observed at this position (Clugston *et al.*, 2003). Therefore, in addition to A domains, C domains act as a second selectivity barrier, which may explain how directionality of nonribosomal peptide synthesis is realized.

1.4.1.4 Thioesterase domain: the nonribosomal hydrolase or cyclase

In order to regenerate the NRPS machinery for the next round of synthesis, the fully assembled peptide has to be promptly released once it reaches the end of the assembly line. This reaction is usually accomplished by the type I thioesterase (TE) domain, which is about 250 amino acids in length and located at the C-terminus of the last module. The first insight into the protein structure of this domain type was gained from the crystal structure of Srf-TE, a TE domain excised from the surfactin synthetases, showing that TE domains belong to the family of α/β hydrolases (Bruner *et al.*, 2002). The full-length peptide chain tethered to the most downstream T domain is first transferred to the hydroxyl group of the highly conserved serine residue within the GxSxG core motif of the TE domain, resulting in the formation a peptidyl-O-TE intermediate (Figure 1.4.6A) (Schneider and Marahiel, 1998). The ester bond is subsequently cleaved either by the attack of a water molecule (hydrolysis) to give a linear peptide carboxylate, or by intramolecular attack of an internal nucleophile (cyclization) to produce a cyclic or branched cyclic peptide (Figure 1.4.6A). The released peptide backbones may undergo postsynthetic modifications, such as cross-linking or halogenation, by the action of dedicated enzymes associated with NRPSs (Walsh *et al.*, 2001). These chemical modifications enlarge the structural diversity of nonribosomal peptides, and introduce structural rigidity and stability against proteolytic degradation. Furthermore, the constrained structures of nonribosomally synthesized peptides ensure precise interaction with their cellular targets.

Since it is difficult to achieve peptide cyclization in non-biological organic chemistry synthesis without protecting all side chains, most of the studies on TE domains have focused on the cyclization mechanism for the development of new strategies to synthesize cyclopeptides. Biochemical characterization of the TE domains excised from several NRPS systems indicated that TE domains display high levels of stereoselectivity and regioselectivity by choosing one specific residue within the peptide chain as the intramolecular nucleophile for cyclization (Grunewald *et al.*, 2004; Kohli *et al.*, 2001; Trauger *et al.*, 2000; Tseng *et al.*, 2002). In addition, it was shown that the peptide chain is pre-organized in a product-like conformation by intramolecular hydrogen bonds during the cyclization reaction, which probably plays an important role in substrate recognition by the TE domain (Trauger *et al.*, 2001). Investigation of the substrate specificity of TE domains would allow generation of diverse cyclic peptide libraries that could be probed for improved or novel biological activities (Kohli *et al.*, 2002).

In some NRPS gene clusters, genes encoding 250-aa proteins with sequence similarity to type II thioesterases were identified. Mutagenesis studies showed that these external thioesterases are important but not essential for effective peptide production (Schneider and Marahiel, 1998). The specific function of type II thioesterases (TEIIs) in nonribosomal synthesis is assumed to be regeneration of inactive NRPSs with acyl groups attached to the 4'-PP cofactors of T domains, the so-called misprimed NRPSs. The posttranslational apo-to-holo modification of T domains by binding of the 4'-PP cofactors is catalyzed by 4'-PP transferases, using CoA as the 4'-PP donor. Due to the low substrate specificity of 4'-PP transferases, various CoA derivatives, mostly present in the form of acyl-CoAs, can serve as substrates in the modification process. However, the acylated 4'-PP cofactors transferred onto T domains would prevent recognition by adjacent domains, resulting in inactive misprimed NRPSs. In these cases, TEIIs can efficiently regenerate the misprimed NRPSs to the active holo form by hydrolyzing these acyl groups (Figure 1.4.4B) (Schwarzer *et al.*, 2003). More recently, it was suggested that TEIIs can also recover activity of NRPSs that are loaded with incorrect amino acid substrates by the A domains and have thus become stalled during peptide synthesis (Yeh *et al.*, 2004). Both findings may explain the observation that disruption of TEII genes caused drastically reduced production of the corresponding peptides, but did not completely abolish the productivity (Schneider and Marahiel, 1998).

1.4.1.5 Modifying domains: tailoring enzymes embedded in NRPSs

In addition to the incorporation of various unusual amino and carboxy acids, peptide synthetases can also introduce structural diversity into the peptide products by carrying out modifications of substrates during peptide chain elongation.

N-methylation of the peptide backbone is a typical modification of nonribosomally synthesized peptides, especially those of fungal origin, which prevents peptide bonds from proteolytic cleavage. This modification is mediated by a 450-aa N-methylation (M) domain that generally resides between the adenylation and thiolation domains (Burmester *et al.*, 1995; Haese *et al.*, 1993; Haese *et al.*, 1994). Transfer of the methyl group from the cofactor S-adenosylmethionine (SAM) to the thioester-bound amino acid probably occurs prior to peptide bond formation (Figure 1.4.7A).

Another common structural feature of many nonribosomal peptides is the occurrence of D-amino acids, which inhibits proteolytic degradation and contributes to the bioactive conformation. In most cases, modules incorporating D-amino acids contain an extra epimerization (E) domain (about 450 residues), located at the C-terminal end of the thiolation domain. The E domain catalyzes racemization of the T domain-bound amino acid or of the C-terminal amino acid of the peptidyl chain, to form an equilibrium between the L- and D-conformers (Figure 1.4.7B). To ensure exclusive incorporation of the D-amino acid into the peptide chain, the C domain within the downstream module selectively accepts the D-enantiomer at the donor site, and catalyzes the condensation reaction between the upstream D-amino acid and the downstream amino acid (Belshaw *et al.*, 1999; Linne and Marahiel, 2000; Luo *et al.*, 2002; Stachelhaus and Walsh, 2000).

Two novel modification domains have been identified in modules responsible for heterocyclic ring formation, such as thiazolines, oxazolines, thiazole and oxazole rings. The 450-aa heterocyclization (Cy) domain, which replaces the usual C domain in these modules, first carries out peptide bond formation and then catalyzes heterocylization of the side chain of serine/threonine/cysteine residues with the newly formed peptide bond

(Figure 1.4.7C) (Gehring *et al.*, 1998; Keating *et al.*, 2000; Miller and Walsh, 2001). The resulting thiazolines or oxazolines can be further oxidized into the aromatic thiazole or oxazole rings by the action of an FMN-dependent oxidation (Ox) domain (about 250 aa, Figure 1.4.7C), which resides C-terminal to either the A or T domain (Du *et al.*, 2001; Schneider *et al.*, 2003; Silakowski *et al.*, 1999).

As an alternative in termination of nonribosomal peptide synthesis, a reductase (R) domain of 350 aa in length, in place of the TE domain, is involved in the biosynthesis of gramicidin A and myxochelin A (Gaitatzis *et al.*, 2001; Kessler *et al.*, 2004; Silakowski *et al.*, 2000). In these systems, the R domain utilizes NADPH as a cofactor to reduce the C-terminal carboxy group of the synthesized peptide chain and release a linear aldehyde or even the corresponding alcohol (Figure 1.4.6B).

1.4.2 Rational design of NRPSs towards construction of novel peptides

The modular peptide synthetases represent the protein templates for nonribosomal peptide synthesis, with each module composed of several structurally and functionally independent domains catalyzing the sequential enzymatic reactions on the growing peptide chains. The nature, number and order of modules and domains within NRPSs determine the primary structure of the peptide product. With the ever-increasing need for new antibiotics and pharmaceutical drugs, the modular organization of NRPSs could be utilized for combinatorial biosynthesis of novel bioactive peptides by rearrangement of modules and domains to construct hybrid peptide synthetases. During manipulation, the choice of the fusion site between modules or domains is one of the most crucial points. Biochemical and structural studies have led to precise identification of the domain borders, as well as the linker regions in between (Mootz *et al.*, 2000). The lack of highly conserved residues within the linkers gives them the advantage of being suitable positions for artificial fusions of domains or modules. A minimal elongation module of NRPSs consists of three core domains in the N- to C- order of C-A-T, suggesting three different strategies to recombine individual domains (Mootz and Marahiel, 1999): (a) intramodular fusion between a C domain and an A domain (C-AT), (b) intramodular fusions between an A domain and a T domain (CA-T), and (c) intermodular fusions between a T domain and a C domain (CAT-CAT) (Figure 1.4.8).

The first type of fusion (Figure 1.4.8a) has been employed for targeted exchange of the A domain and the accompanying T domain in the Leu-activating module of the surfactin synthetases, for several other AT di-domains of bacterial or fungal origins with different amino acid specificities (Stachelhaus *et al.*, 1995). These engineered synthetases produced the predicted surfactin variants, but at a much lower level compared with the native system, which could be explained by the substrate specificity of C domains at the acceptor site for the cognate amino acid.

The second intramodular fusion between an A domain and a T domain (Figure 1.4.8b) was expected to overcome the limitation imposed by the C domain's selectivity. One successful example of this type of fusion has been reported in construction of bi-modular hybrid enzymes in vitro (Doekel and Marahiel, 2000). In these experiments, two different A domains, one from the tyrocidine synthetase TycB and the other from the bacitracin synthetase BacA, were separately fused to a T domain that is attached to the termination module (including a TE domain) of the tyrocidine synthetase TycC. Production of the predicted dipeptides by the hybrid NRPSs was observed, suggesting that the interaction between foreign A and T domains is efficient enough to allow thiolation of noncognate amino acids. This result also indicates a considerable degree of tolerance towards noncognate substrates at the donor site of C domains, in contrast to the specificity at the acceptor site.

The third possible fusion site is located between two complete modules (Figure 1.4.8c). The feasibility of artificial module fusion was first demonstrated by generation of di-modular hybrid peptide synthetases through fusion the Pro-activating module with the Orn-activating module, or with the Leu-activating module, all from the tyrocidine synthetases (Mootz *et al.*, 2000). The TE domain was also fused to the C-terminus of the di-modular peptide synthetases. Subsequently, incubation of the hybrid NRPSs with the D-Phe-incorporating module of the tyrocidine synthetase TycA, gave rise to the predicted tripeptides D-Phe-Pro-Orn, and D-Phe-Pro-Leu.

The strategies presented above have focused on the recombination of NRPS modules or domains to rebuild a new biosynthetic system for a desired product. In comparison, targeted modification of natural peptides by site-directed mutagenesis of the specificity-conferring residues within the A domains (Challis *et al.*, 2000; Eppelmann *et al.*, 2002; Stachelhaus *et al.*, 1999), is a rather small alteration (Figure 1.4.8d). For example, in the study of CDA, the calcium-dependent lipopeptide antibiotics produced by *Streptomyces coelicolor*, the substrate specificity of the seventh Asp-activating module of CDA synthetases was rationally changed to asparagine recognition by point mutations of two residues in the corresponding A domain, which led to production of a CDA analog containing Asn residue, instead of Asp, at the seventh position (Uguru *et al.*, 2004). However, one drawback of this approach was the dramatic reduction in productivity of the engineered synthetases, which could be caused by the substrate specificities of the downstream C domain and the terminal TE domain.

In addition, there are many other manipulations of practical interest to generate structural diversity in the peptide products, such as introduction or deletion of modification domains at a certain position (Figure 1.4.8e), manipulation of ring size in cyclic peptides, or construction of hybrid products by fusion of NRPSs with modular polyketide synthases (Section 1.6). It has also been possible to alter the length of the peptide chain by translocation of the C-terminal TE domain (Figure 1.4.8f) (de Ferra *et al.*, 1997).

1.5 Nonribosomally synthesized lipopeptides

Lipopeptides are composed of a linear or cyclic peptide with a long fatty acid moiety covalently attached to its N-terminus, and they have received considerable attention for their antimicrobial and surfactant properties, as well as other physiological activities. Lipopeptides are produced by a variety of bacterial genera, including *Bacillus*, *Pseudomonas*, and *Streptomyces*, through nonribosomal peptide synthetases. It has been suggested that the fatty acid moieties of lipopeptides are incorporated into the peptide products by direct transfer of the fatty acid to the first activated amino acid of the peptide chain. Although enzymes catalyzing this reaction have not been identified so far, a condensation domain preceding the first NRPS module was observed in all reported lipopeptide synthetases, serving as a good candidate for this catalytic activity.

Bacillus subtilis strains are known to produce two groups of cyclic lipopeptides: the surfactin group, including surfactin, lichenysin, fengycin (synonymous to plipastatin), esperin, which contains a β-hydroxy fatty acid, and the iturin group, including iturin, mycosubtilin, bacillomycin, which contains a β-amino fatty acid. The peptide synthetase operons for iturin-type lipopeptides have been completely sequenced and characterized (Duitman *et al.*, 1999; Moyne *et al.*, 2004; Tsuge *et al.*, 2001). The deduced peptide synthetase enzymes share the same modular organization as other NRPS systems, but additionally contain functional domains with homology to fatty acid synthases/polyketide synthases (Section 1.6) and an amino transferase domain (presumably responsible for activation of a long-chain fatty acid, modification with a β-amino group, and incorporation of the resulting β-amino fatty acid), along with the first module of NRPSs to form a hybrid synthetase (Figure 1.5.1A). In contrast, genes required for synthesis and/or modification, and incorporation of the β-hydroxy fatty acid moieties of surfactin-type lipopeptides, have not yet been identified in either the peptide synthetase operons or the flanking regions (Figure 1.5.1B) (Cosmina *et al.*, 1993; Konz and Marahiel, 1999; Steller *et al.*, 1999; Tosato *et al.*, 1997). The lipohexapeptide fusaricidin produced by *P. polymyxa* is structurally similar to surfactin in that both contain a β-hydroxy fatty acid. However, unlike surfactin-type lipopeptides, the lipid moiety of fusaricidin is modified by addition of a guanidino group and thus it is reasonably postulated that the fusaricidin biosynthetic gene cluster may harbor genes encoding enzymes involved in fatty acid synthesis and/or modification, as seen in the mycosubtilin biosynthetic operon (Figure 1.5.1A).

1.6 Hybrid peptide-polyketide synthetase systems

Polyketides are a large group of natural products synthesized through sequential condensation of simple carboxylic acids, and some of known polyketides have been widely used in medicine and agriculture for their antagonistic activities against bacteria, fungi, viruses, and parasites, as well as for their immunosuppressive and antitumor properties. Synthesis of the macrolide subgroup of polyketides (e.g. erythromycin) is catalyzed by large multifunctional enzyme complexes known as Type I polyketide synthases (PKSs) with modular organizations resembling those of nonribosomal peptide synthetases (Cane and Walsh, 1999). A minimal elongation module of PKSs also consists of three enzymatic domains: an acyl transferase (AT) domain for selection and transfer of carboxylic acid monomers, an acyl carrier protein (ACP) for docking of the monomers or the growing polyketide chain as enzyme-bound thioesters, and a ketosynthase (KS) domain for decarboxylative condensation between two adjacent acyl thioesters to extend the polyketide chain (Figure 1.6.1). Like T domains (or peptidyl carrier protein, PCP) of NRPSs, ACPs of PKSs are posttranslationally modified by covalent attachment of CoA-derived 4'-phosphopantetheine (4'-PP) groups, catalyzed by a subgroup of 4'-PP transferases specific for the ACPs of fatty acid synthases and PKSs (Walsh *et al.*, 1997). Once the polyketide chain reaches its full length at the most downstream ACP, a specialized thioesterase (TE) domain, usually found at the C-terminus of PKSs, catalyzes the release of the polyketide product. Additional domains responsible for modifications of the carbon backbones of polyketides have also been identified (Figure 1.6.1) (Cane and Walsh, 1999). Together, PKSs apparently share very similar modular organizations and biosynthetic strategies with NRPSs, which allows production of hybrid peptide-polyketide metabolites with novel biological activities by combining individual NRPS and PKS modules for combinatorial biosynthesis, or more directly by engineering naturally occurring hybrid NRPS-PKS synthetases that integrate peptide synthesis and polyketide synthesis.

A rapidly increasing number of hybrid NRPS-PKS systems has been identified and characterized in the past few years. Leinamycin, as one example among many, is an antitumor antibiotic produced by *Streptomyces atroolivaceus*, consisting of an unusual 1,3-dioxo-1,2-dithiolane moiety fused to a hybrid peptide-polyketide macrocyclic ring containing two amino acid residues, D-Ala and L-Cys (Figure 1.6.2). Cloning and sequencing of the leinamycin biosynthetic gene cluster revealed a putative hybrid NRPS-PKS mega-synthetase, which is composed of an NRPS initiation module and an NRPS elongation module, followed by eight PKS modules with unusual domain organization (Figure 1.6.2) (Tang *et al.*, 2004). It was proposed that leinamycin biosynthesis was initiated by activation and loading of the amino acids D-Ala and L-Cys to their cognate NRPS modules. Subsequent cyclization and oxidation of these two residues gave rise to an NRPS-bound thiazole intermediate, which was then transferred to the downstream PKS modules for stepwise elongations with carboxylic acids. The fully assembled hybrid peptide-polyketide backbone was released and cyclized with the action of the C-terminal TE domain to produce a macrolactam intermediate, which was further modified by fusion of the 1,3-dioxo-1,2-dithiolane moiety to yield the mature leinamycin molecule.

Sequence analyses of hybrid NRPS-PKS synthetase genes and comparisons with the known NRPS and PKS genes, have revealed the presence of putative intermodular and interpolypeptide linker regions that were presumably required for transfer of the peptidyl or polyketide chain between NRPS and PKS modules within hybrid NRPS-PKS systems

(Du *et al.*, 2001). These results provided the first mechanistic insight into intermodular communications within hybrid NRPS-PKS synthetases, which must be taken into consideration for future attempt at construction of hybrid peptide-polyketide biosynthetic templates.

1.7 Main objectives of this thesis

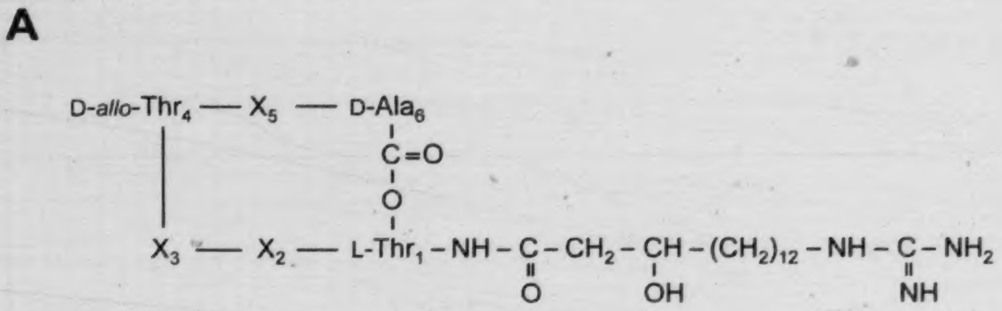
In a previous study by Beatty and Jensen (2002), an environmental bacterium isolated from canola stubble, which was highly inhibitory to the growth of *L. maculans* in vitro, was determined to be a strain of *Paenibacillus polymyxa* based on its biochemical and morphological characteristics, as well as 16S rRNA sequence similarity, and was given the strain designation PKB1. The antifungal activity of PKB1 against *L. maculans* was found to be mainly attributable to production of a mixture of fusaricidin variants A and B around the onset of sporulation (Figure 1.3.1B). Although purification methods and primary structures of fusaricidins have been described, the biosynthetic details of these cyclic depsipeptides are still unknown. The usefullness of *P. polymyxa* PKB1 as a biocontrol agent for the blackleg disease of canola caused by *L. maculans* may be improved by increasing the production of fusaricidin variants associated with the antifungal activity. In order to fulfill this goal, investigation of the detailed biosynthetic steps involved in fusaricidin production is a necessary first step. On the basis of their structural similarity to several lipopeptides isolated from *Bacillus* strains (e.g. surfactin and lichenysin), it was hypothesized that the peptide moiety of fusaricidin is synthesized nonribosomally by peptide synthetases. The study presented in this thesis was initiated by isolation of the peptide synthetase gene responsible for fusaricidin biosynthesis in *P.*

polymyxa PKB1. The reverse genetics method established by Turgay and Marahiel (1994), using degenerate primers derived from the highly conserved motifs within NRPSs, was adopted to amplify peptide synthetase gene fragments from the PKB1 genome, which were subsequently used to identify the putative gene encoding fusaricidin synthetase by probing the genomic DNA libraries. Mutagenesis studies were carried out to determine if the putative fusaricidin synthetase gene is involved in the production of fusaricidins by the PKB1 strain. In addition, the chromosomal regions surrounding the fusaricidin synthetase gene were analyzed to identify the complete fusaricidin biosynthetic gene cluster. Finally, the amino acid specificity of an internal adenylation domain within the fusaricidin synthetase was investigated in vitro.[1]

25

---

[1] *A version of this chapter has been published. Li et al. 2007. Applied and Environmental Microbiology. 73: 3480-3489.*


A
D-allo-Thr4
X5
D-Ala6
C=O
O
X3
X2
L-Thr1 - NH - C - CH2 - CH - (CH2)12 - NH - C - NH2
O
OH
NH

X_2: D-Val, D-*allo*-Ile, or D-Ile

X_3: L-Tyr, L-Val, L-Ile, L-Phe, or L-*allo*-Ile

X_5: D-Asn, or D-Gln

B

Fusaricidin	Amino acid positions		
	X_2	X_3	X_5
A	D-Val	L-Val	D-Asn
B	D-Val	L-Val	D-Gln

Figure 1.3.1 Primary structure of the fusaricidin-type lipopeptide antibiotics produced by *P. polymyxa*. (A) Amino acid substitutions tolerated at three defined positions in fusaricidin variants are presented. Residues are numbered according to the order of synthesis. (B) Amino acid substitutions in fusaricidins A and B, the two variants associated with the antifungal activity of PKB1 strain against *L. maculans*.

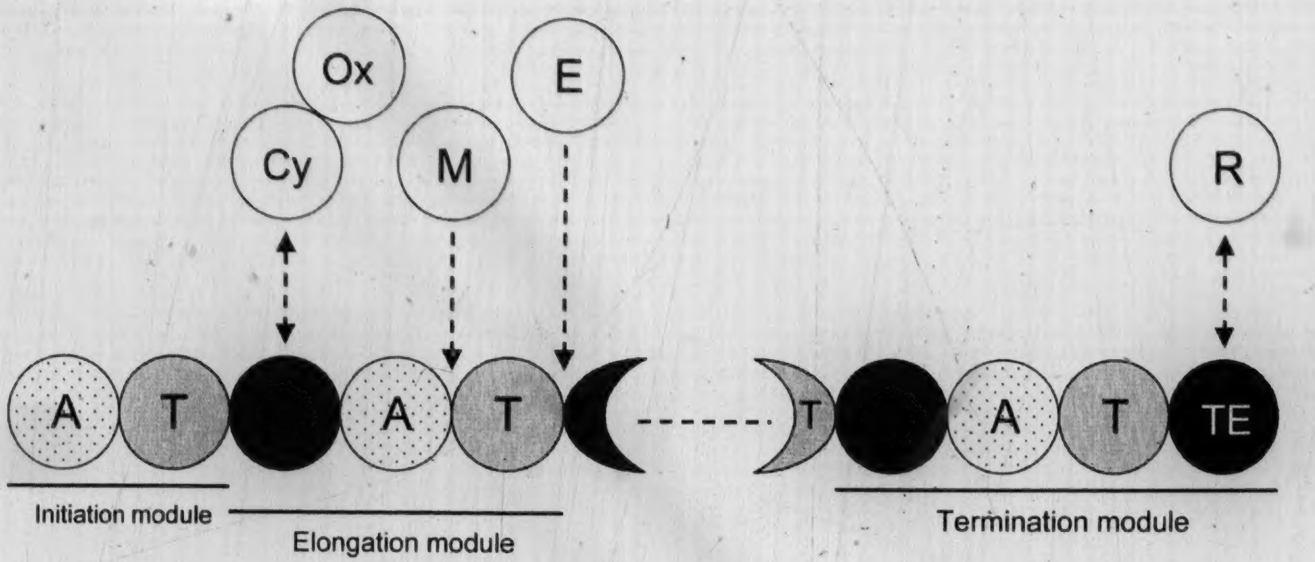


Figure 1.4.1 Schematic representation of the module and domain organization of nonribosomal peptide synthetases. A minimal elongation module typically consists of three core domains: an adenylation (A) domain, a thiolation (T) domain, and a condensation (C) domain. The initiation module usually lacks an N-terminal C domain, and the termination module contains an additional thioesterase (TE) domain at the C-terminus. In the case of heterocyclic ring formation, the C domain is replaced by a heterocyclization (Cy) domain, which is often coupled to an oxidation (Ox) domain. The N-methylation (M) domain and the epimerization (E) domain are optional domains that introduce modifications of the amino acids being incorporated into the elongating peptidyl chain. The TE domain is sometimes replaced by a reductase (R) domain as an alternative mechanism of termination. Adapted from Challis and Naismith (2004).

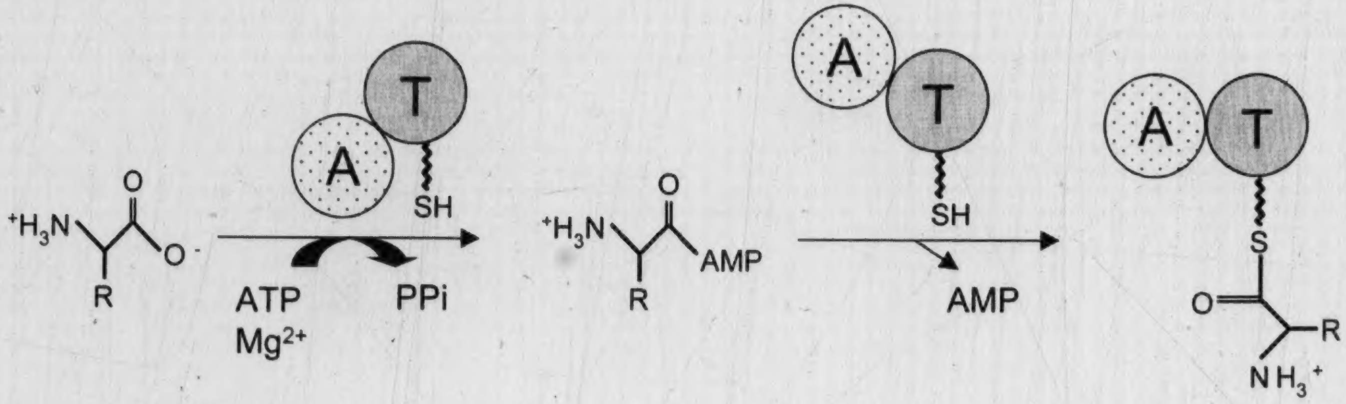


Figure 1.4.2 Schematic diagram of amino acid adenylation and subsequent thiolation during nonribosomal peptide synthesis. The amino acid substrate is first activated as an aminoacyl adenylate by the action of the adenylation (A) domain. This unstable intermediate is then bound as a thioester to the thiol group of the 4'-phosphopantetheine cofactor (represented by a wavy line), which is covalently attached to the thiolation (T) domain located downstream of the A domain.

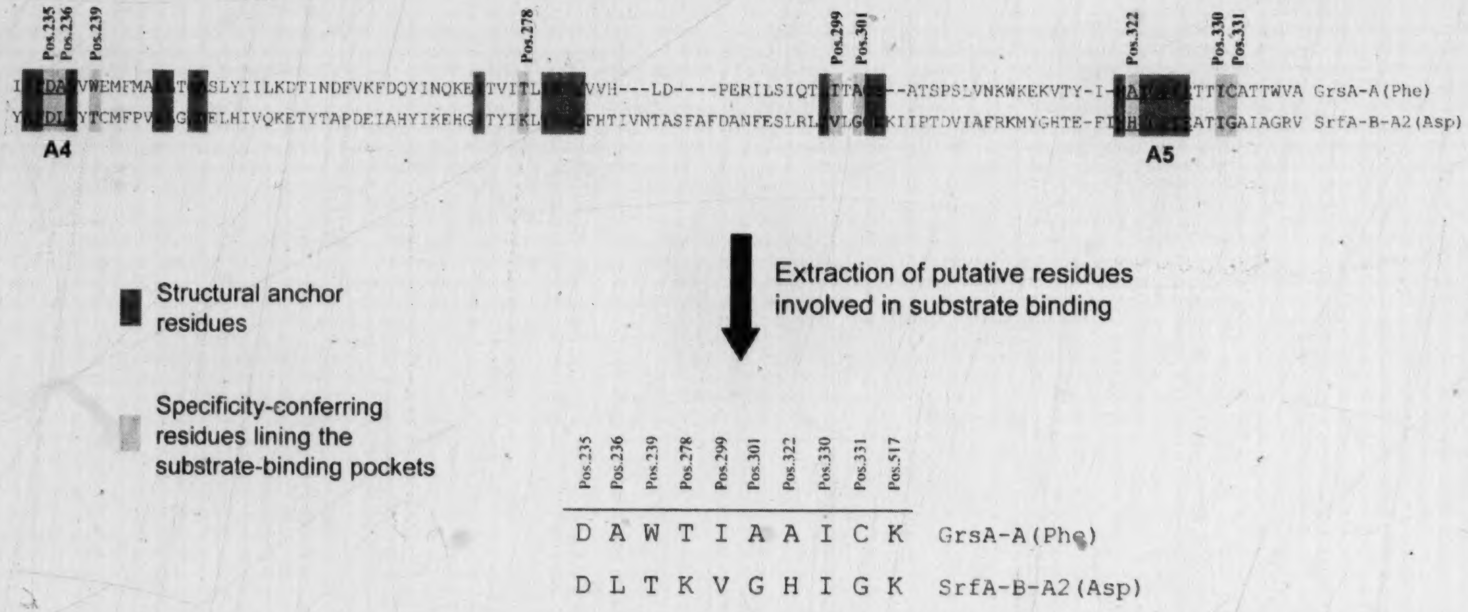

Pos.235
Pos.236
Pos.239
Pos.278
Pos.299
Pos.301
Pos.322
Pos.330
Pos.331
GrsA-A(Phe)
SrfA-B-A2(Asp)
A4
A5
Structural anchor residues
Specificity-conferring residues lining the substrate-binding pockets
Extraction of putative residues involved in substrate binding
Pos.517
D A W T I A A I C K GrsA-A(Phe)
D L T K V G H I G K SrfA-B-A2(Asp)

Figure 1.4.3 Extraction of the substrate specificity-conferring residues by alignment of the amino acid sequences in the region of core motifs A4 and A5. Two examples are shown, the Phe-activating adenylation domain from the gramicidin synthetase GrsA, and the Asp-activating adenylation domain from the surfactin synthetase SrfA-B. The highly conserved position 517 (Lys) is located in the core motif A10. Conserved motif sequences A4 and A5 are underlined. Adapted from Stachelhaus *et al.* (1999).

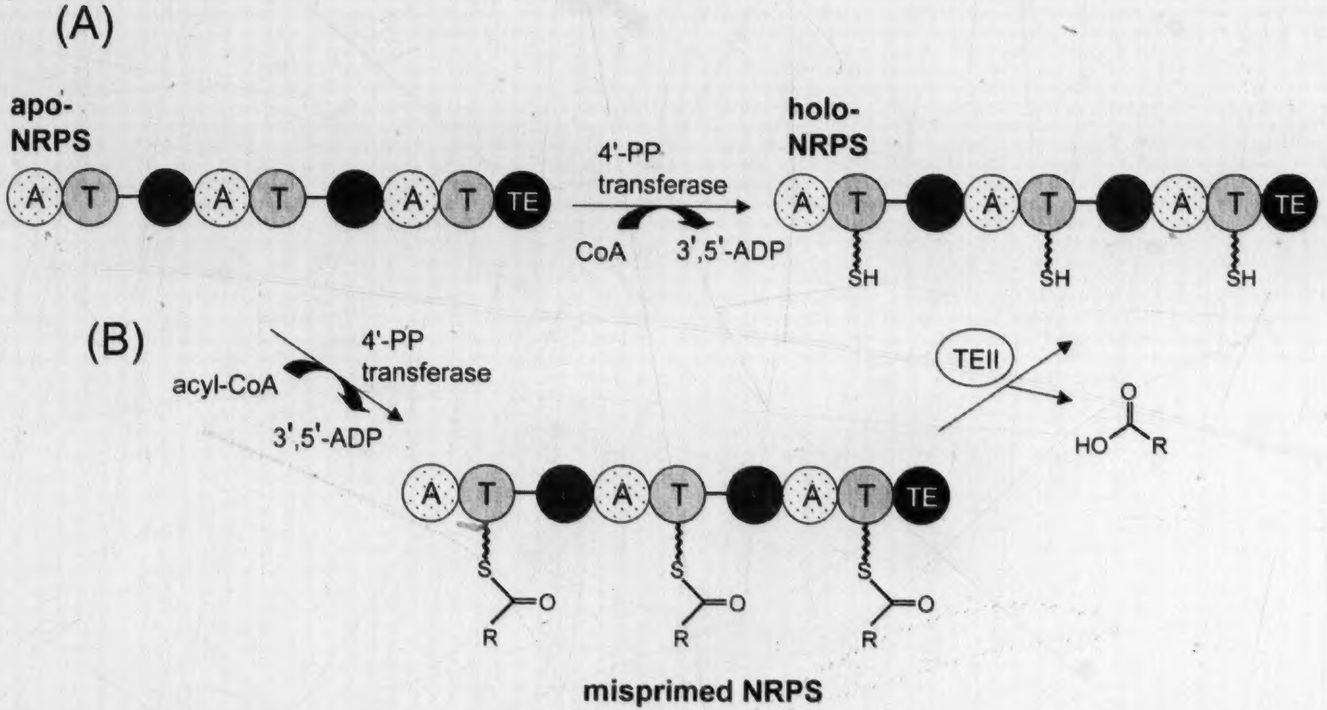

(A)
apo-
NRPS
A
T
A
T
A
T
TE
4'-PP
transferase
CoA
3',5'-ADP
holo-
NRPS
SH
(B)
acyl-CoA
4'-PP
transferase
3',5'-ADP
TEII
HO
O
R
S
misprimed NRPS

Figure 1.4.4 Posttranslational modification of the peptide synthetase (NRPS), from the inactive apo-form to the active holo-form, prior to peptide biosynthesis. (A) The 4'-phosphopantetheine (4'-PP) cofactors, represented by wavy lines, are transferred from coenzyme A (CoA) onto the thiolation (T) domains with the action of dedicated 4'-PP transferases. (B) If acyl-CoAs are used as substrates by the 4'-PP transferases, type II thioesterases (TEIIs) can hydrolyze the acyl groups attached to the 4'-PP cofactors, and then regenerate the misprimed NRPS to its active holo form (Section 1.4.1.4). Adapted from Schwarzer *et al.* (2002).

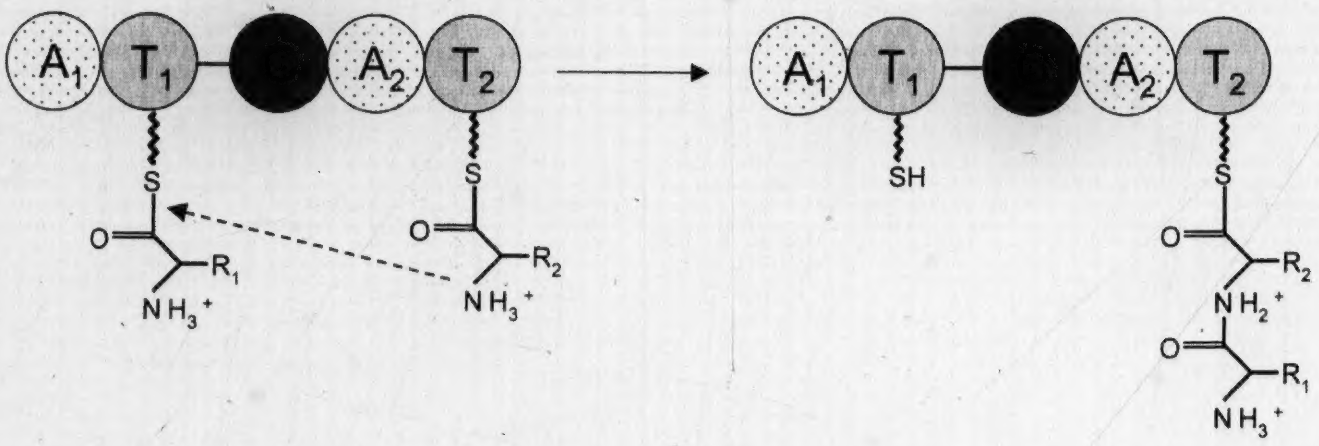

A_1
T_1
A_2
T_2
S
O
R_1
NH_3^+
S
O
R_2
NH_3^+
A_1
T_1
SH
A_2
T_2
S
O
R_2
NH_2^+
O
R_1
NH_3^+

Figure 1.4.5 Peptide bond formation between two amino acids that are thioester-bound to the thiolation (T) domains of two successive modules, mediated by the condensation (C) domain. Shown is the nucleophilic attack of the α-amino group of the downstream substrate on the electrophilic thioester bond of the upstream substrate to give an amide bond during peptide biosynthesis.

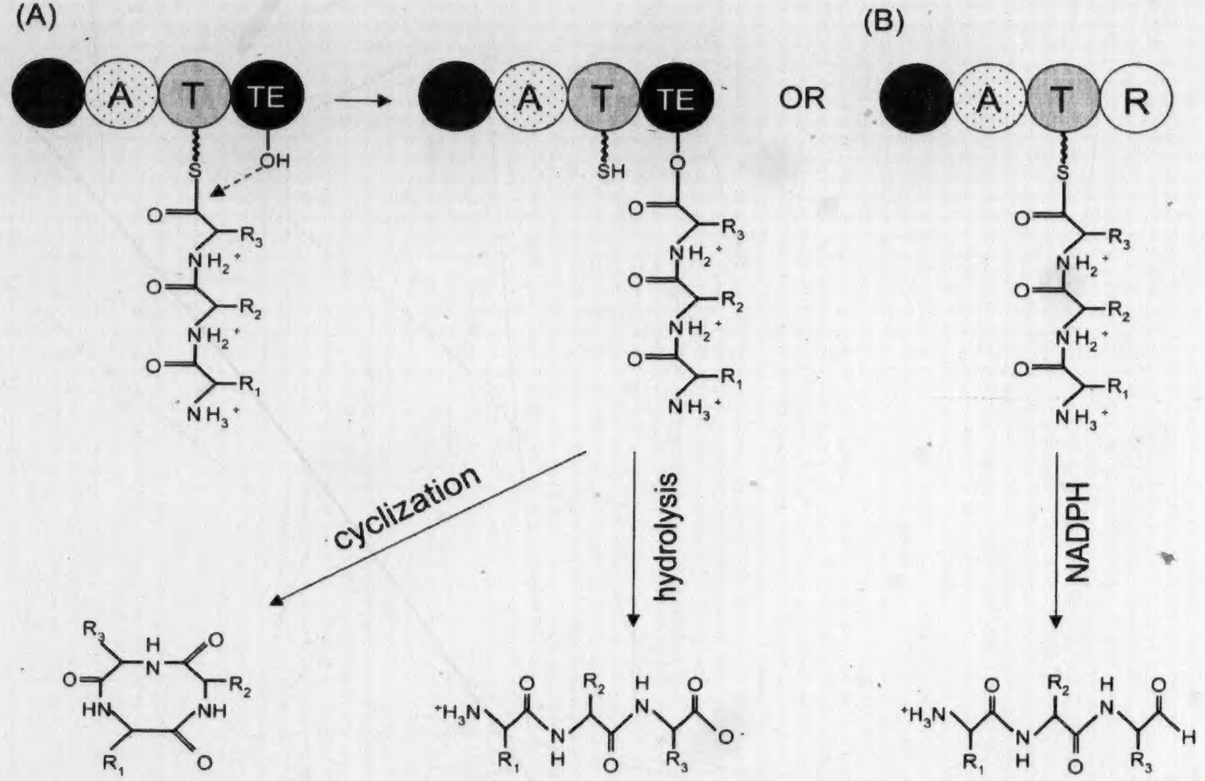
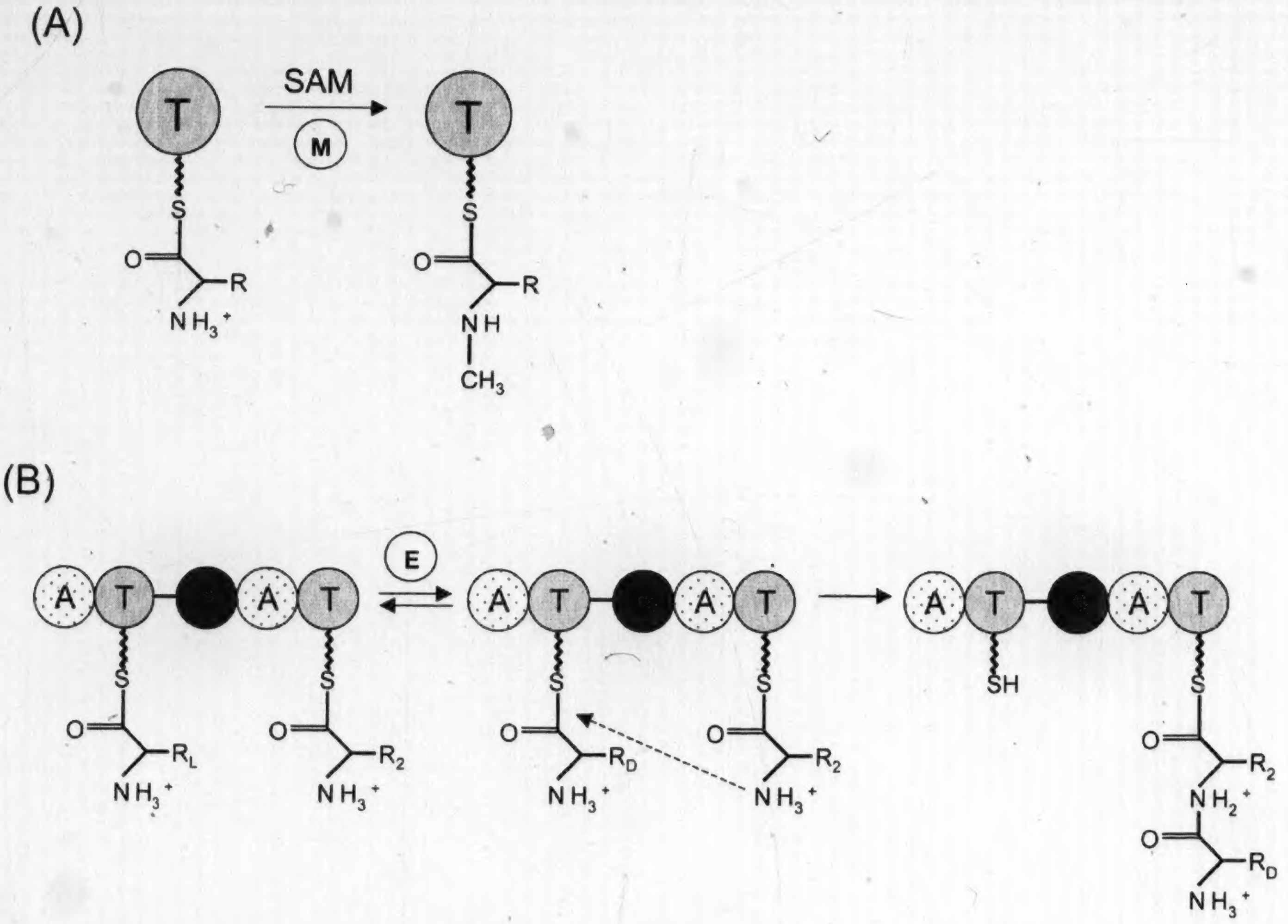

(A)
(B)
A
T
TE
R
OR
S
OH
SH
NH_2^+
NH_3^+
R_1
R_2
R_3
cyclization
hydrolysis
NADPH
^+H_3N

Figure 1.4.6 A schematic diagram showing two termination strategies of nonribosomal peptide synthesis. (A) Hydrolysis or cyclization of the peptide chain catalyzed by the C-terminal thioesterase (TE) domain. (B) Alternatively, the C-terminal carboxyl group of the peptide is reduced by the reductase (R) domain to yield an aldehyde.


(A)
SAM
M
T
S
O
R
NH3+
NH
CH3
(B)
E
A
T
S
O
RL
R2
RD
NH3+
SH
NH2+

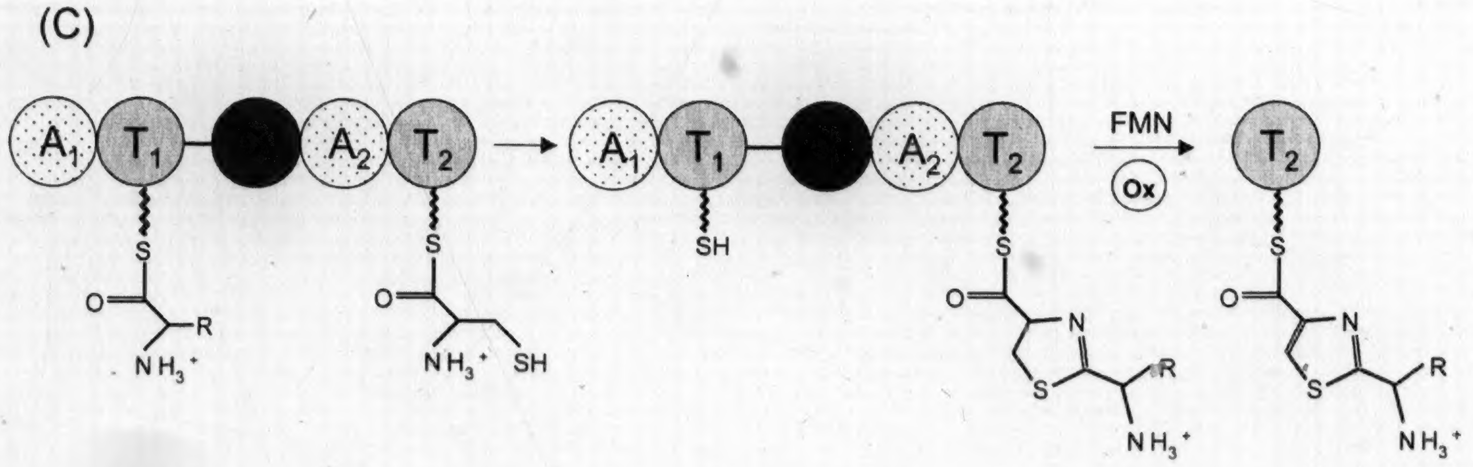


Figure 1.4.7 Modifications of amino acid substrates being incorporated into nonribosomal peptides, catalyzed by optional modifying domains. (A) The N-methylation (M) domain mediates the transfer of the methyl group from the cofactor S-adenosylmethionine (SAM) to the amino group of a thioester-bound amino acid. (B) The epimerization (E) domain catalyzes the racemization of the thioester-bound amino acid to form an equilibrium between the L- and D-conformers, and subsequently the condensation (C) domain exclusively incorporates the D-conformer into the peptide chain. (C) The heterocyclization (Cy) domain catalyzes both peptide bond formation and heterocyclization of the side chain of cysteine (or serine, or threonine) with the amino acid bound to the preceding module. The heterocyclic rings may be further oxidized by the action of an additional oxidation (Ox) domain, using FMN (flavin mononucleotide) as the cofactor. Adapted from Konz and Marahiel (1999).

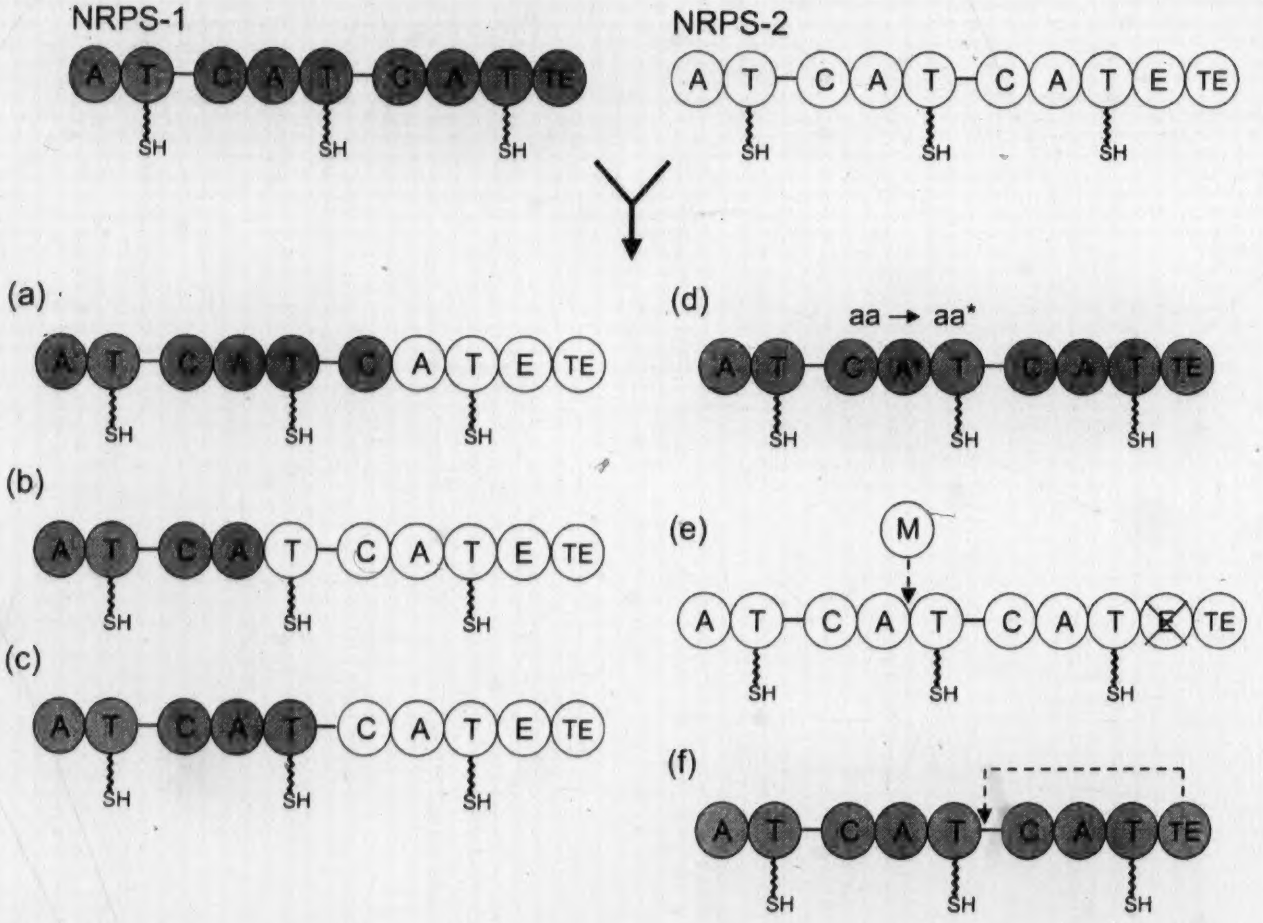

NRPS-1
A T C A T C A T TE
SH SH SH
NRPS-2
A T C A T C A T E TE
SH SH SH
(a)
A T C A T C A T E TE
(b)
A T C A T C A T E TE
(c)
A T C A T C A T E TE
(d)
aa → aa*
A T C A* T C A T TE
(e)
M
A T C A T C A T E TE
(f)
A T C A T C A T TE

Figure 1.4.8 Possible strategies to construct hybrid peptide synthetases for the synthesis of new peptide products.

Figure 1.4.8 (continued) Recombination of peptide synthetase templates can be realized by domain fusions (a) between condensation (C) domains and adenylation (A) domains, (b) between adenylation (A) domains and thiolation (T) domains, or (c) between thiolation (T) domains and condensation (C) domains. Other strategies include (d) alteration of the substrate specificity of A domains (aa→aa*), (e) introducing modifying domains (such as an N-methylation (M) domain) or deleting a modifying domain (such as an epimerization (E) domain), (f) translocation of the thioesterase (TE) domain to shorten the length of the peptide product. Adapted from Mootz and Marahiel (1999).

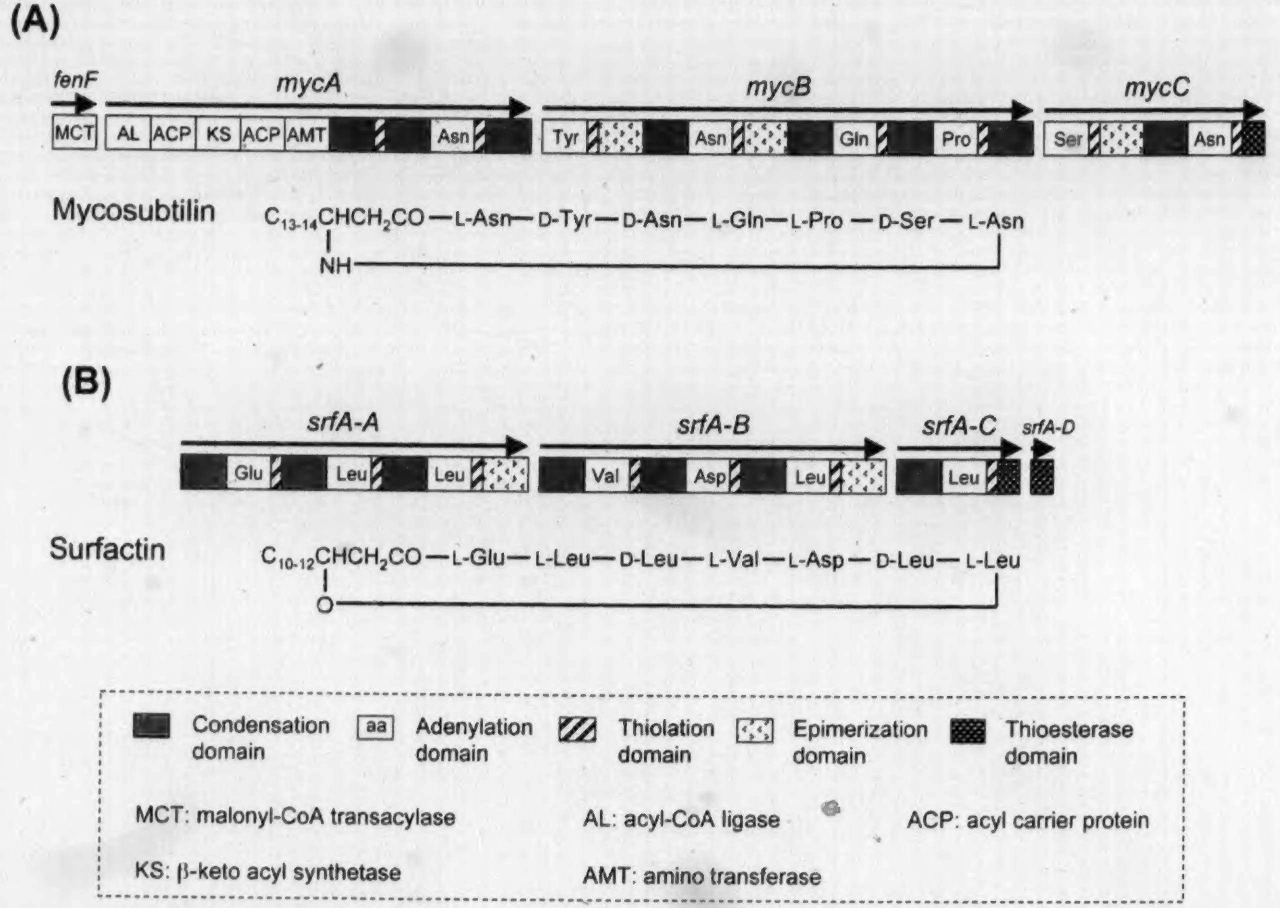

(A)
fenF
mycA
mycB
mycC
MCT
AL
ACP
KS
ACP
AMT
Asn
Tyr
Asn
Gln
Pro
Ser
Asn
Mycosubtilin
C13-14CHCH2CO — L-Asn — D-Tyr — D-Asn — L-Gln — L-Pro — D-Ser — L-Asn
NH
(B)
srfA-A
srfA-B
srfA-C
srfA-D
Glu
Leu
Leu
Val
Asp
Leu
Leu
Surfactin
C10-12CHCH2CO — L-Glu — L-Leu — D-Leu — L-Val — L-Asp — D-Leu — L-Leu
O
Condensation domain
aa
Adenylation domain
Thiolation domain
Epimerization domain
Thioesterase domain
MCT: malonyl-CoA transacylase
AL: acyl-CoA ligase
ACP: acyl carrier protein
KS: β-keto acyl synthetase
AMT: amino transferase

Figure 1.5.1 Nonribosomally synthesized cyclic lipopeptides produced by *Bacillus subtilis.*

Figure 1.5.1 (continued) Two examples are shown, (A) mycosubtilin carrying a β-amino fatty acid, and (B) surfactin carrying a β-hydroxy fatty acid. The ORF arrangement of each lipopeptide biosynthetic operon, the deduced domain organizations of the peptide synthetases, and the primary structure of the produced lipopeptide are illustrated. The amino acid substrates, all in L-configuration, of the adenylation domains are indicated.

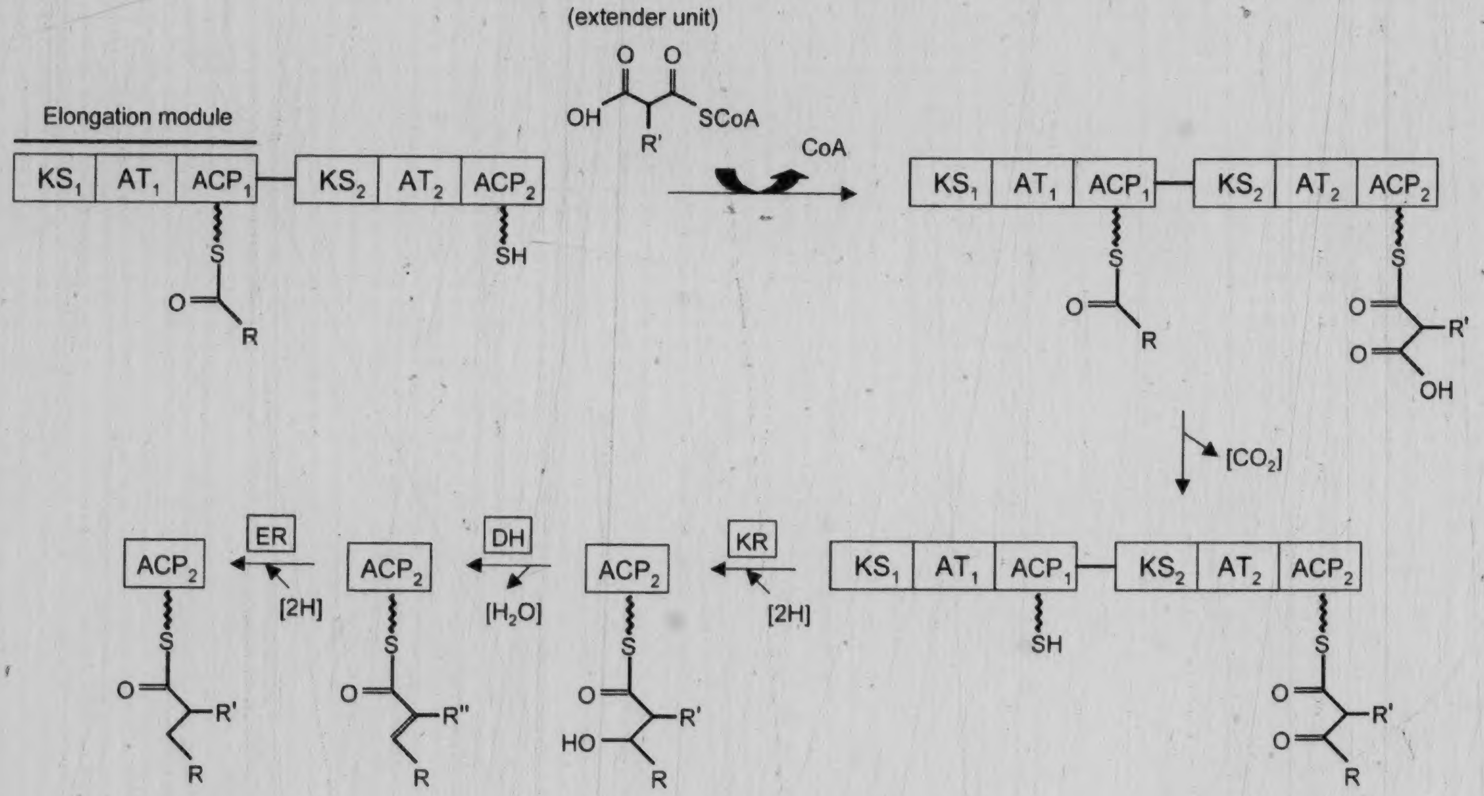

(extender unit)
Elongation module
KS₁
AT₁
ACP₁
KS₂
AT₂
ACP₂
SH
O
R
OH
SCoA
R'
CoA
[CO₂]
KR
[2H]
DH
[H₂O]
ER
[2H]
HO
R''

Figure 1.6.1 Schematic diagram of catalytic functions of Type I PKS domains.

Figure 1.6.1 (continued) A minimal elongation module of Type I PKSs consists of three core domains: an acyl transferase (AT) domain, an acyl carrier protein (ACP), and a ketosynthase (KS) domain. The AT domain selects appropriate monomer substrate, usually malonyl-CoA (R'=H) or methylmalonyl-CoA (R'=Methyl) as the extender unit, and transfers the already activated acyl group to the thiol group of the 4'-phosphopantetheine cofactor (represented by a wavy line) covalently attached to the adjacent ACP. The KS domain then catalyzes decarboxylative condensation between the upstream acyl group of the growing polyketide chain and the downstream thioester-bound extender unit to elongate and translocate the polyketide chain to the downstream ACP. Prior to the next condensation cycle, the β-carbonyl group may remain unchanged, or be partially or fully reduced with the action of three optional modifying domains: first, ketoreduction by ketoreductase domain (KR) to yield a β-hydroxyl group, followed by dehydration (by dehydratase domain, DH) to introduce the α,β unsaturation in the polyketide chain, and finally, enoylreduction (by enoyl reductase domain, ER) to produce a methylene function at the β-carbon.

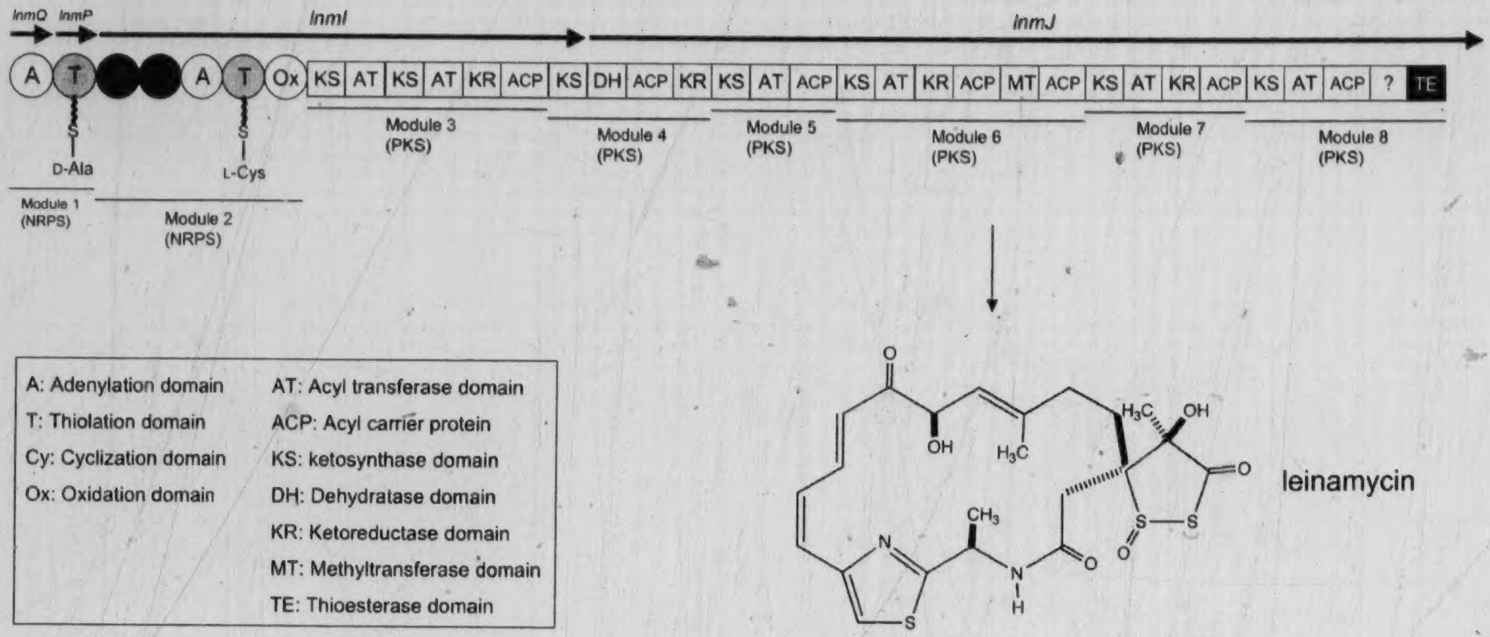

lnmQ
lnmP
lnmI
lnmJ
A
T
A
T
Ox
KS
AT
KS
AT
KR
ACP
KS
DH
ACP
KR
KS
AT
ACP
KS
AT
KR
ACP
MT
ACP
KS
AT
KR
ACP
KS
AT
ACP
?
TE
S
D-Ala
S
L-Cys
Module 1 (NRPS)
Module 2 (NRPS)
Module 3 (PKS)
Module 4 (PKS)
Module 5 (PKS)
Module 6 (PKS)
Module 7 (PKS)
Module 8 (PKS)
A: Adenylation domain
T: Thiolation domain
Cy: Cyclization domain
Ox: Oxidation domain
AT: Acyl transferase domain
ACP: Acyl carrier protein
KS: ketosynthase domain
DH: Dehydratase domain
KR: Ketoreductase domain
MT: Methyltransferase domain
TE: Thioesterase domain
O
OH
H3C
H3C
OH
O
CH3
N
S
N
H
O
S
S
O
leinamycin

Figure 1.6.2 Proposed modular organization of the hybrid NRPS-PKS synthetase encoded by the leinamycin biosynthetic gene cluster (*lnm*). Genes encoding NRPSs, hybrid NRPS-PKS, and PKSs are represented by black arrows. The amino acid residues to be incorporated into leinamycin are shown as NRPS-bound thioesters. The question mark represents a domain of unknown function. The methyltransferase (MT) domain catalyzes O-methylation of the β-hydroxyl group (Figure 1.6.1). Adapted from Tang *et al.* (2004).

Table 1.1 The conserved core motifs found in the common catalytic domains of nonribosomal peptide synthetases. Adapted from Konz and Marahiel (1999) and Schwarzer *et al.* (2003).

Domain	Core motif [a]	Consensus Sequence [b]
Adenylation	A1	L(TS)YxEL
	A2 (core 1)	LKAGxAYL(VL)P(LI)D
	A3 (core 2)	LAYxxYTSG(ST)TGxPKG
	A4	FDxS
	A5	NxYGPTE
	A6 (core3)	GELxIxGxG(VL)ARGYL
	A7 (core4)	Y(RK)TGDL
	A8 (core5)	GRxDxQVKIRGxRIELGEIE
	A9	LPxYM(IV)P
	A10	NGK(VL)DR
Thiolation	T (core6)	DxFFxxLGG(HD)S(LI)
Condensation	C1	SxAQxR(LM)(WY)xL
	C2	RHExLRTxF
	C3 (His)	MHHxISDG(WV)S
	C4	YxD(FY)AVW
	C5	(IV)GxFVNT(QL)(CA)xR
	C6	(HN)QD(YV)PFE
	C7	RDxSRNPL
Thioesterase	TE	GxSxG
N-methylation	M1 (SAM)	VL(DE)GxGxG
	M2	NELSxYRYxAV
	M3	VExSxAROxGxLD
Epimerization	E1	PIOxWF
	E2 (His)	HHxISDG(WV)S
	E3 (race A)	DxLLxAxG
	E4 (race B)	EGHGRE
	E5 (race C)	RTVGWFTxxYP(YV)PFE
	E6	PxxGxGYG
	E7 (race D)	FNYLG(QR)

[a] Former nomenclature is given in parentheses.

[b] Single-letter amino acid codes are used. Alternative amino acids in a specific position are indicated in parentheses; X represents any amino acid at that position.

2. Materials and methods

2.1 Strains of bacteria and fungi, plasmids, reagents and supplies

The strains and plasmids used in this study are listed in Table 2.1 and Table 2.2, respectively. *P. polymyxa* PKB1 (=ATCC 202127) and *Leptosphaeria maculans* were obtained from P. Kharbanda, Alberta Research Council (Vegreville, Canada). The *Staphylococcus* plasmid vector pC194, which replicates in *P. polymyxa* and imparts chloramphenicol resistance, and the *E. coli* plasmids pBEST501 and pBEST502, carrying the same neomycin resistance gene selectable in *Bacillus subtilis*, were obtained from the Bacillus Genetic Stock Center (Ohio State University, Columbus, OH). The *E. coli* plasmid pBM-SPECT carrying a spectinomycin resistance marker functional in *Bacillus subtilis* was provided by S.L. Wong, Department of Biological Sciences, University of Calgary. Cosmid and fosmid clones identified from the PKB1 genomic DNA libraries, or prepared for mutagenesis studies, are described in Section 3 (Results and discussion) along with their use.

Chemicals used in preparations of media and buffers were reagent grade, and were purchased from Sigma (St. Louis, MO), BD (Sparks, MD), or ICN (Aurora, OH) unless otherwise indicated. Restriction endonucleases and other DNA modifying enzymes were obtained from Roche (Laval, QC) and New England Biolabs (Mississauga, ON). Both electrophoresis apparatus and electroporation apparatus were purchased from Bio-Rad Laboratories (Hercules, CA). Hybond-N nylon membranes used for Southern analyses and colony hybridizations were purchased from Amersham Biosciences (Buckinghamshire, UK). DNA oligonucleotide primers were synthesized by Molecular Biology Service Unit (MBSU, Department of Biological Sciences, University of

Alberta), Qiagen (Alameda, CA), or IDT (Coralville, IA). An Eppendorf 5415 microcentrifuge was used for microcentrifugation, and a Beckman Model J2-21 centrifuge was used for centrifugation of large-volume bacterial cultures. Bacterial and fungal cultures were grown using a cell production roller drum (Bellco Biotechnology), or a G-24 incubator shaker (New Brunswick Scientific CO.). The sources of all the other materials used in this study are described in the subsequent subsections along with their specific use.

2.2 *E. coli* procedures

2.2.1 Growth conditions

Luria-Bertani (LB) medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl), or SOB medium (2% tryptone, 0.5% yeast extract, 0.05% NaCl, 2.5 mM KCl, pH 7.0, supplemented with 10 mM $MgCl_2$ and 20 mM $MgSO_4$ before use), was commonly used to grow *E. coli* cultures at 37°C unless otherwise indicated. When required, antibiotics were added to growth media at the following concentrations: ampicillin, 100 μg/ml; apramycin, 50 μg/ml; chloramphenicol, 25 μg/ml; kanamycin, 50 μg/ml; neomycin, 50 μg/ml; streptomycin, 50 μg/ml; spectinomycin, 50 μg/ml. In addition to the antibiotic (12.5 μg/ml chloramphenicol), LB medium used for isolation of fosmid DNA from *E. coli* cultures was supplemented with 1x Replication Induction Solution (L-arabinose, Lucigen, Middleton, WI), which induces the expression of a medium-copy origin of replication (*ori*V) in the vector and thus increases the fosmid yield. All *E. coli* strains were stored at -70°C as 20% glycerol stocks.

2.2.2 Preparation of *E. coli* electrocompetent cells and electroporation

The desired *E. coli* strain was grown for 16 h in 2 ml LB medium at 37°C, from which 1 ml of the fresh 16 h culture was used to inoculate 200 ml of LB medium in a 500 ml flask. The culture was then incubated at 37°C, shaken at 250 rpm until the OD_{600} reached 0.5-0.6. The cells were harvested by centrifugation at 2,500 x g for 10-15 min at 4°C, washed once with 200 ml, then with 100 ml, and then with 50 ml of 4°C 10% glycerol. The washed cells were resuspended in 3-4 ml of 4°C 10% glycerol per 100 ml of the original culture. During these steps, care was taken to keep the cells chilled at 4°C at all times. The cell suspension was then flash frozen in a dry ice-ethanol bath in 40 µl aliquots and stored at -70°C for future use.

E. coli BW25113/pIJ790 and BW25113/pKD46 electrocompetent cells used in the PCR-targeting mutagenesis studies were prepared using SOB medium, which was incubated at 28°C to maintain the temperature-sensitive plasmids. When necessary, L-arabinose (10 mM) was added to the growth medium to induce the expression of λ Red recombination functions. Cells harvested from the 50 ml original culture (with an OD_{600} of ~0.6) were washed once with 50 ml and then once with 25 ml of 4°C 10% glycerol. After the last wash, the supernatant was discarded and the cell pellet was resuspended in the residual 100-150 µl 10% glycerol, and 50 µl of the cell suspension was used immediately for electroporation.

For transformation of DNA into *E. coli*, previously frozen electrocompetent cells were slowly thawed on ice before use. Plasmid DNA or ligation product was added to a microcentrifuge tube containing 40 µl electrocompetent cells. The mixture was transferred into a 2 mm gap disposable electroporation cuvette (Molecular BioProducts) placed in a Bio-Rad Gene Pulser® II apparatus, and pulsed with an electric charge of 2.5 kV and 25 μF capacitance. Immediately after the pulse, 1 ml of 4°C LB medium was added to the cuvette, and the cells were then allowed to recover by incubation in a cell production roller drum at 37°C or 28°C for 1-2 h. To isolate the desired transformants, the transformation mixture was plated on LB agar (1.5% agar) containing appropriated antibiotic(s), and the plates were incubated for 16-24 h at either 37°C or 28°C as required. In some cases, 40 μl of 100 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and 80 μl of 2% X-gal (5-bromo-4-chloro-3-β-D-galactopyranoside) were spread on each plate before inoculation for blue-white selection.

2.2.3 Preparation of colony lifts

The cosmid/fosmid-containing *E. coli* colonies were patched with sterile toothpicks onto LB agar plates containing appropriate antibiotics. The plates were incubated at 37°C for 14-16 h, and then chilled at 4°C for at least 30 min before use. To take the colony lift, a Hybond-N nylon membrane of the appropriate size was carefully placed on the agar surface, and its position was marked on the bottom side of the plate. After 1 min, the membrane was gently removed and treated with a series of solutions, in order to lyse the cells and release the cosmid/fosmid DNA. The membrane was placed with the colony side up on 3MM paper filters (Whatman, Philadelphia, PA) saturated with (1) 10% SDS for 1 min, (2) denaturation solution (0.5 M NaOH, 1.5 M NaCl) for 5 min, (3) neutralization solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.5) for 3 min and this step was repeated once more. Finally, the membrane was vigorously washed in 2x SSC buffer (0.3 M NaCl, 0.03 M tri-sodium citrate) to remove the cell debris, and air dried for 1 h. DNA was fixed onto the membrane by baking for 2 h at 80°C in a vacuum oven, and then the membrane was wrapped in plastic and stored at room temperature.

2.3 *Paenibacillus polymyxa* procedures

2.3.1 Growth conditions and media

P. polymyxa was routinely cultivated in glucose broth (GB, 1% glucose, 1% peptone, 0.2% beef extract, 0.1% yeast extract, 0.5% NaCl) or on GB agar plates (containing 1.5% agar) at 37°C for fast vegetative growth (Rosado and Seldin, 1993). Fusaricidin production by *P. polymyxa* was optimized at 28°C in a complex medium PDB-soy, a mixture of equal volumes of potato dextrose broth (PDB, Difco) and Soy medium (1.5% soy bean flour, 4.7% soluble starch, 0.01% KH_2PO_4, 0.02% $FeSO_4$-$7H_2O$, pH 6.8) (Salowe *et al.*, 1990). *P. polymyxa* culture grown in TCSS liquid medium (3% trypticase soy, 1% soluble starch) at 37°C for 24 h was used to inoculate PDB-soy medium as 1% (v/v) inoculum, which was then incubated at 28°C while being shaken at 250 rpm. Brain Heart Infusion (BHI) medium was used for conjugation experiments and for the preparation of *P. polymyxa* genomic DNA and plasmids. When required, antibiotics were added to growth media at the following concentrations: apramycin, 25 µg/ml; chloramphenicol, 5-10 µg/ml; polymyxin B, 25 µg/ml. Potato dextrose agar (PDA, 2.4% PDB, 1.5% agar) was used for growth of *L. maculans* and for the fusaricidin bioassay. *Paenibacillus* spore stocks were prepared by scraping the sporulated colonies from one PDA plate into 0.5-1 ml 20% glycerol in one 1.5 ml microcentrifuge tube, and the spore stocks were stored at -70°C.

2.3.2 Purification of genomic DNA from *P. polymyxa*

Genomic DNA from *P. polymyxa* was prepared by a modified version of the protocol used for isolation of genomic DNA from *Streptomyces* (Hopwood *et al*. 1985). *P. polymyxa* culture, grown in 50 ml BHI for 16-20 h at 28°C and shaken at 250 rpm, was harvested by centrifugation at 12,000 x g for 10 min. The cell pellet was washed twice with 10.3% sucrose, and then resuspended in 4 ml fresh lysozyme solution (25 mM Tris-HCl, 25 mM EDTA, 0.3 M sucrose, pH 8.0) containing 4 mg/ml lysozyme and 100 µg/ml RNase A. The sample was incubated at 37°C for 45 min, followed by the addition of SDS and proteinase K to give final concentrations of 2% and 0.2 mg/ml, respectively. The cell lysate was incubated for 15 min at 37°C, then extracted twice with equal volumes of phenol-chloroform-isoamyl alcohol solution (25:24:1, v/v) and then twice with equal volumes of chloroform-isoamyl alcohol solution (24:1, v/v). After the last extraction, genomic DNA in the aqueous phase was precipitated by adding two volumes of 95% ethanol and 0.1 volume of 3 M sodium acetate. The tube was gently inverted until a white fibrous material came out of the solution, and then kept on ice for 5-10 min. The clump of genomic DNA was wound up using a sterile micropipette tip and washed once in 70 % ethanol. After centrifugation for 5 min, the DNA pellet was air dried at room temperature for 10 min, dissolved in 50-200 µl of 10 mM Tris-HCl buffer (pH 8.5), and stored at 4°C.

In some cases for PCR reactions, Sigma® GenElute™ Bacterial Genomic DNA Kit was used for quick preparation of high-purity genomic DNA from 2 ml BHI cultures grown at 37°C for 14-16 h.

2.3.3 Plasmid isolation from *P. polymyxa*

Plasmid DNA was isolated from *P. polymyxa* using a method previously described by Belliveau and Trevors (1989) with a few modifications. Briefly, the plasmid-bearing *P. polymyxa* strain was grown at 37°C for 16-18 h in 10 ml BHI medium supplemented with the appropriate antibiotic(s) using a cell production roller drum. The culture was split into six 1.5 ml microcentrifuge tubes, and the cells were collected by centrifugation at 14,000 rpm for 5 min. The cell pellet in each tube was resuspended in 100 µl of E buffer (40 mM Tris-HCl, 2 mM EDTA, 6.7% sucrose, pH 8.0) supplemented with 2 mg/ml lysozyme, and the tubes were incubated at 37°C for 30 min. After adding 200 µl fresh lysis buffer (50 mM Tris-HCl, pH 8.0, 6.7% sucrose, 3% SDS, 0.15 M NaOH), the tubes were gently inverted 20 times to mix the contents, followed by incubation at 60°C for 20 min. The cell lysates were neutralized with 100 µl of 2 M Tris-HCl (pH 7.0), mixed by gentle inversion, and incubated at 37°C for 20 min. Following the addition of 50 µl of 5 M NaCl, the tubes were kept on ice for an additional 1 h before centrifugation. The supernatants containing plasmid DNA were transferred to fresh tubes, and treated with RNase A (100 µg/ml) and proteinase K (100 µg/ml) for 15 min at 37°C. Then the samples were extracted once with equal volumes of phenol-chloroform-isoamyl alcohol solution and once with equal volumes of chloroform-isoamyl alcohol solution, after which plasmid DNA in the aqueous phase was precipitated with 2 volumes of 98% ethanol for 16-24 h at -70°C. The DNA pellets were obtained by centrifugation at maximum speed for 15 min, washed once with 70% ethanol, and redissolved in 10 µl of 10 mM Tris-HCl buffer (pH 8.5). The DNA samples from six preparations were pooled into one microcentrifuge tube and stored at -20°C.

2.3.4 Preparation of *P. polymyxa* electrocompetent cells and electroporation

The wild-type *P. polymyxa* PKB1 strain was grown in 20 ml GB liquid medium at 28°C for 24 h, shaken at 250 rpm. This starter culture (2 ml) was used to inoculate 200 ml GB broth in a 500 ml flask, which was then incubated at 28°C, shaken at 250 rpm, until the OD_{600} reached ~0.7-0.8. The cells were harvested by centrifugation at 10,000 x g for 15 min at 4°C, washed twice in 50 ml 4°C sterile distilled water and then once in 50 ml 4°C HEB buffer (272 mM sucrose, 1 mM $MgCl_2$, 7 mM HEPES, pH 7.3) (Rosado and Seldin, 1993). After the washes, the cells were resuspended in 2 ml 4°C HEB buffer, and 800 μl aliquots of the cell suspension were dispensed into 1.5 ml microcentrifuge tubes. The cell aliquots were then used immediately for electroporation, or flash frozen in a dry ice-ethanol bath and stored at -70°C for future use.

For electroporation of plasmid DNA into *P. polymyxa*, previously frozen electrocompetent cells were slowly thawed in an ice-water bath, and then the DNA sample (0.5 μg) was added to one microcentrifuge tube containing 800 μl electrocompetent cells. The contents were mixed by gently pipetting, and transferred into a 4 mm gap disposable electroporation cuvette (Molecular BioProducts). After incubation on ice for 5 min, the cells were pulsed with an electric charge of 2.5 kV and 25μF capacitance, using a Bio-Rad Gene Pulser® II apparatus. Immediately after the pulse, 1 ml 4°C GB broth was added to the cuvette, and the cell suspension was kept on ice for 15 min before transferring to a sterile 15 ml disposable centrifuge tube containing 7 ml GB liquid medium. The electroporated cells were recovered by growing at 28°C for 2 h, shaken at 100 rpm, after which the cells were collected by centrifugation at 3,000 x g for 5 min. The cell pellet was resuspended in 500 μl GB broth, and aliquots of 100 μl were spread on GB plates containing the appropriate antibiotic(s), followed by incubation at 37°C for up to 24 h.

2.3.5 Transfer of DNA into *P. polymyxa* by intergeneric conjugation

Both *E. coli* donor strain (ET12567 containing the non-transmissible plasmid pUZ8002 and the plasmid/cosmid/fosmid to be mobilized) and the wild-type PKB1 recipient strain were grown in BHI broth (supplemented with antibiotics as appropriate) for 12-14 h at 37°C, shaken at 200 rpm. Donor and recipient cells were washed three times with equal volumes of fresh BHI medium, and then mated (donor / recipient ratio = 5:1) on 0.22 μm Millipore® filters (Millipore Corporation, Bedford, MD) placed on BHI plates. Following 16 h of incubation at 28°C, cells from each filter were resuspended in 2 ml of BHI broth and 200 μl aliquots were spread on GB agar plates containing chloramphenicol (5 μg/ml) to select for *P. polymyxa* exconjugants, and polymyxin B (25 μg/ml) to counterselect against the *E. coli* donor strain. Usually, chloramphenicol resistant colonies that appeared after 24 h incubation at 37°C were patched onto fresh GB plates containing higher concentrations of chloramphenicol (10 μg/ml) and polymyxin B (50 μg/ml) for a clearer selection of chloramphenicol resistant *P. polymyxa* exconjugants.

2.3.6 Antifungal activity bioassay

Production of fusaricidins by *P. polymyxa* strains was detected by bioassay with *L. maculans* as the indicator organism. *L. maculans* was stored at -70°C as a 20% glycerol stock of pycnidiospores, which was prepared by scraping the lawn of *L. maculans* grown on PDA plates into sterile 20% glycerol. The mixture was then vigorously vortexed and filtered through sterile cotton fluff and aliquots of 1 ml were collected into microcentrifuge tubes. For the fusaricidin bioassay, PDA plates (24.5 cm x 24.5 cm square petri dishes) were spread with 0.5-1 ml of *L. maculans* spore stock, and sample wells were punched out of the agar with a sterile cork borer. For small-scale extraction of antifungal material, *P. polymyxa* was grown in 20 ml PDB-soy medium at 28°C for 72 h, shaken at 250 rpm. Culture samples (1.5 ml) were harvested by microcentrifugation at 14,000 rpm for 10 min, and cell pellets were resuspended in 100 μl of methanol with the aid of a small sterile plastic pestle. Following incubation at room temperature for 30 min, cell suspensions were centrifuged and the fusaricidin-containing supernatants were applied to wells of the bioassay plate, along with the solvent control (100 μl of methanol). Alternatively, 750-μl amounts of whole cultures, or culture supernatants, or uninoculated PDB-soy media (the control) were mixed with equal volumes of methanol and supplemented with 10 μl of 1:10 diluted glacial acetic acid. After 30 min incubation at room temperature, cells were harvested by centrifugation and 500 μl of each supernatant was air dried to ~100 μl. These concentrated extracts were then applied to wells of the bioassay plate. Bioassay plates were incubated for 24 h at room temperature exposed to room light, and then at 28°C in the dark for an additional 72 h.

2.3.7 HPLC analysis of fusaricidin production

P. polymyxa PKB1 or *fusA* mutant strain A4 were grown in 200 ml PDB-soy medium at 28°C for 48 h. Each culture was mixed with an equal volume of methanol and then supplemented with KCl at a concentration of 5% (w/v). After 30 min of incubation at room temperature, each cell suspension was harvested by centrifugation, and 25 ml of the supernatant was applied to a SepPak C18 cartridge (Waters) that had been prewetted with 5 ml of methanol and rinsed with 5 ml of water. The loaded cartridge was then eluted with 5 ml of water, 5 ml of 40% methanol, and finally 5 ml of 80% methanol. The 80% methanol extracts were concentrated to dryness under a stream of air and dissolved in 1.0 ml of 50% methanol. Samples (0.05 ml) of each 80% methanol concentrate were analyzed by C18 reversed-phase chromatography on a Bondclone 10 μ C18 column (8 by 100 mm; Phenomenex) at a flow rate of 2.0 ml/min using a 15-min linear gradient ranging from 30% methanol to 90% methanol, both in 0.1% formic acid, with detection at 220 nm. Fractions (0.5 ml) were collected across the gradient, dried, and bioassayed against *L. maculans*.

2.4 DNA manipulations

Routine DNA manipulations and analyses were carried out as described by Sambrook *et al.* (1989). Restriction digestion of DNA was performed according to the manufacturers' protocols (New England Biolabs, or Roche). The ligation reaction was 20 μl in volume, and the molar ratio of the insert to the vector is maintained at 3:1-5:1. The mixture of vector, insert and sterile distilled water was pre-heated at 50°C for 5 min and then kept on ice for 2 min. After adding 1 μl T4 DNA ligase (Roche) and the corresponding 1x ligation buffer, the ligation mixture was incubated at 16°C for 16-20 h. In the case of blund-end ligation, PEG 8000 was added into the reaction to give a final concentration of 15% (w/v).

2.4.1 Gel purification of DNA fragments

DNA fragments fractionated by electrophoresis on standard agarose gels were purified using the QIAquick Gel Extraction Kit (Qiagen, Mississauga, ON) following the manufacture's instruction. DNA was eluted from the spin column into 30 µl elution buffer supplied with the kit.

P. polymyxa genomic DNA fragments were also purified from 0.8% low-melting-point agarose gels by agarase digestion. Briefly, the gel slices containing DNA fragments of desired sizes were excised and soaked in two volumes (1 mg ~ 1 µl) of 1x β-Agarase buffer (New England Biolabs) in a 1.5 ml microcentrifuge tube. The tube was incubated at 4°C for 30 min, and then heated at 65°C until the agarose gel completely dissolved (~10 min). After the molten agarose was cooled to ~42°C, the enzyme β-Agarase I (New England Biolabs) was added to the sample at 1 unit per 100 mg of the original agarose gel. The tube was incubated at 42°C for an additional 3 h to allow complete digestion of agarose. Following the addition of 0.1 volume of 3 M sodium acetate, the tube was chilled at 4°C for 15 min, microcentrifuged at 14,000 rpm for 15 min, and the supernatant was transferred to a new microcentrifuge tube. DNA was precipitated by adding 2 volumes of isopropanol along with incubation for up to 24 h at -20°C. Finally, the DNA pellet was washed in 70% isopropanol, air dried and redissolved in an appropriate volume of TE buffer (pH 8.0).

2.4.2 Polymerase chain reaction (PCR)

Routine PCR reactions (25 µl) contained 100 ng plasmid DNA or 250 ng genomic DNA as template, 10 pmol of each primer, and 0.5 µl (~2.5 U) Taq DNA polymerase (Fermentation Service Unit, Department of Biological Sciences, University of Alberta) in

1x Taq buffer (50 mM Tris-HCl, pH 9.0, 1.5 mM $MgCl_2$, 0.4 mM β-mercaptoethanol, 0.1 mg/ml bovine serum albumin, 10 mM ammonium sulfate, 0.2 mM of each dNTP). PCR reactions were carried out using thin-walled microcentrifuge tubes (0.2 ml) along with a T-Gradient thermocycler (Biometra, Goettingen, Germany). The oligonucleotide primers used for PCRs described in this thesis are listed in Table 2.3.

For preliminary identification of *P. polymyxa* mutants, colony PCR was performed using fresh GB cultures grown at 37°C for 16 h. Cells harvested from 100 µl of *P. polymyxa* culture were washed once with 100 µl of 10.3% sucrose solution, resuspended in 25 µl sterile distilled water, and heated at 95°C for 15 min. Following microcentrifugation at 7,000 rpm for 5 min, 5-10 µl of the supernatant was used as template in a 25 µl PCR reaction.

For amplification of a DNA fragment corresponding to the internal adenylation domain (FusA-A6) of the fusaricidin synthetase, the Expand High Fidelity PCR system (Roche) was used to improve the fidelity of the PCR product. Amplification was conducted in a 50 µl reaction containing 250 ng fosmid DNA (6D11), 300 nM of each primer, 200 µM of each dNTP, and 0.75 µl (2.6 U) enzyme mix in 1x Buffer 2 (with 1.5 mM $MgCl_2$), under the following PCR conditions: 94°C for 2 min; 10 cycles of 94°C for 15 s, 55°C for 30 s, and 72°C for 90 s; 20 cycles of 94°C for 15 s, 55°C for 30 s, and 72°C for 90 s with the extension time increased by 5 s per cycle; and a final elongation step of 72°C for 7 min.

For generation of long PCR products, PCR reactions were set up using the Expand Long Template PCR system (Roche). The 50 µl reaction mixture consisted of 500 ng genomic DNA, 500 nM of each primer, 500 µM of each dNTP, and 0.75 µl (3.75 U) enzyme mix in 1x Buffer 2 (containing 2.75 mM $MgCl_2$). Amplification was carried out with the following program: 94°C for 2 min; 10 cycles of 94°C for 10 s, 55°C for 30 s, and 68°C for 7 min; 20 cycles of 94°C for 15 s, 58°C for 30 s, and 68°C for 7 min with the extension time increased by 20 s per cycle; and a final elongation step of 68°C for 7 min.

For direct cloning of PCR-amplified DNA fragments, 1 unit of Taq DNA polymerase was added to the reaction mixture immediately after amplification. The tube was incubated at 72°C for 10 min in order to introduce the 3' A-overhangs necessary for cloning into the pCR®2.1-TOPO® vector (Invitrogen, Carlsbad, CA). The TOPO® cloning reaction was set up following the manufacturer's instructions.

2.4.3 Transfer of DNA fragments onto nylon membranes

Restriction enzyme digested DNA was fractionated by agarose gel electrophoresis, stained in ethidium bromide solution, and photographed under UV light. The agarose gel was treated by soaking in depurination solution (0.25 M HCl) for 10 min with gentle agitation, then in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 30 min with gentle agitation, and finally in neutralization buffer (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.5) for 30 min with gentle agitation. Fractionated DNA fragments were transferred onto a Hybond-N nylon membrane (Amersham) by capillary blotting following the manufacturer's instructions. DNA was fixed onto the membrane by baking for 2 h at 80°C in a vacumm oven. The membrane was either used immediately in hybridization experiments or wrapped in plastic and stored at room temperature in a dark place for future use.

2.4.4 Labeling of DNA probes

Double-stranded DNA fragments or linearized plasmids were labeled with [α-^{32}P]dCTP by nick translation (Hopwood *et al.*, 1985; Sambrook *et al.*, 1989). The reaction mixture (30 µl) contained 0.1-1 µg probe DNA, 1 µl each of 1 mM unlabelled dATP, dGTP and dTTP (Roche), 1 µl (10 µCi) [α-^{32}P]dCTP (Amersham), 2.5 µl of 1:10^4 diluted DNAse I (Roche), and 0.5 µl DNA polymerase I (Roche) in 1x nick translation buffer (50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 1 mM DTT, 50 µg/ml BSA). After incubation at 15°C for 1-2 h, the nick translation reaction was terminated by adding 15 µl of 0.5 M EDTA (pH 8.0). The radiolabeled DNA probe was then denatured by boiling for 5 min, chilled on ice briefly, and used immediately in hybridization experiments.

2.4.5 Hybridization of Southern and colony blots

The nylon membrane with bound DNA was placed in a hybridization bottle (Robbins Scientific), and was pre-wetted first in sterile distilled water and then in the hybridization buffer (1 mM EDTA, 7% SDS, 0.5 M Na_2HPO_4, pH 7.2). When there was significant overlap of the blot in the tube, a nylon mesh of larger size was used in order to separate the overlapped layers. The membrane was prehybridized with 100-150 µg/ml denatured salmon sperm DNA in 25-50 ml hybridization buffer. The bottle was incubated in a hybridization oven (Fisher Scientific) at 65°C for 1-2 h before the denatured, radiolabeled DNA probe was added. After hybridization for at least 12 h at 65°C, the membrane was rinsed briefly in 2x SSC + 0.1% SDS, and then washed twice (5 min each) in the same solution at room temperature, twice (10 min each) in 1x SSC + 0.1% SDS at 65°C, and three times (5 min each) in 0.1x SSC + 0.1% SDS at 65°C. The membrane was wrapped in plastic, and placed in a phosphorimager (Molecular Dynamics model 445 SI) to develop autoradiograms from 2 h to 24 h.

For re-hybridization analysis, the membrane was stripped of the bound DNA probe by soaking in boiling 0.1% SDS solution. The SDS bath was allowed to cool to room temperature, and the stripping procedure was repeated once. Finally, the membrane was rinsed with 2x SSC, and the removal of the radiolabeled probe was confirmed using the phosphorimager.

2.4.6 DNA sequencing and analysis

Routine DNA sequencing was performed using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham). The reaction (20 µl) was set up with 500 ng plasmid DNA, or 800-1000 ng of cosmid/fosmid DNA partially digested with *Sca*I, as template. The reaction mixture also contained 5 pmol of primer and 4-8 µl of the ET sequencing premix in 1x ET buffer (80 mM Tris-HCl, 2 mM $MgCl_2$, pH 9.0). The sequencing reaction was carried out in a Biometra® T-Gradient thermocycler using the program consisting of 30 cycles of 96°C for 30 s, 60°C for 1 min. After the reaction cycles, 2 µl salt solution (1.5 M sodium acetate, pH>8.0, 250 nm EDTA) and 80 µl of 95% ethanol were added to the tubes, which were then kept on ice for 15 min followed by centrifugation at room temperature for 15 min. The DNA pellet was washed once with 70% ethanol, air dried for up to 5 min, and then sent to MBSU at University of Alberta, from where the DNA sequence information was obtained. The nucleotide sequences of cosmids Col-8 and Col-19 were determined by SeqWright DNA Technology Services (Houston, TX). DNA sequencing of two fosmids, 4G9 and 6D11, was carried out in part by SeqWright, and gaps were closed by direct sequencing of fosmid DNA, performed by MBSU. The nucleotide sequence data were compiled and analyzed using either GeneTools 2.0 (BioTools Inc., Edmonton, Canada) or DNA Strider 1.2 (Marck, 1988). The online programs BLAST (http://www. ncbi.nlm.nih.gov/BLAST/) and ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) were used for homology searches and prediction of open reading frames (ORFs), respectively.

2.4.7 Preparation of *P. polymyxa* genomic libraries

The first *P. polymyxa* genomic DNA library was constructed using the SuperCos-1 Cosmid Vector Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. Briefly, the SuperCos-1 vector was digested with *Xba*I, dephosphorylated with calf intestine alkaline phosphatase (CIAP), and then digested with *Bam*HI. Genomic DNA from *P. polymyxa* PKB1 was partially digested with *Sau*3AI for an appropriate time interval to obtain insert fragments predominantly of the desired size (30-42 kb). After dephosphorylation with CIAP, the insert DNA was ligated into the *Bam*HI site of the pre-cut cosmid vector. The resulting ligation mixture was then packaged *in vitro* with Gigapack III Gold packaging extract (Stratagene), and transfected into *E. coli* XL1-Blue MR host strain supplied along with the kit.

The second SuperCos-1 library was constructed following the same procedures mentioned above, except that the *Sau*3AI partially digested genomic DNA from PKB1 strain was size-fractionated by sucrose gradient centrifugation before ligation into the cosmid vector. A sucrose solution in sterile distilled water (22.5%) was frozen in a centrifuge tube for at least 12 h at -20°C in a vertical position and thawed slowly at room temperature. This freeze-thaw cycle was repeated once more, and care was taken not to disturb the gradient (10%-35%) that had formed. The DNA sample was gently loaded onto the surface of the sucrose gradient, and then the tube was centrifuged using a SW-40 Ti rotor at 35,000 rpm for 24 h at 4°C, after which fractions of approximately 700 µl were transferred into clean tubes. The contents of each fraction were examined by agarose gel electrophoresis to determine which fractions contain genomic DNA fragments between 30 kb and 42 kb. The DNA fragments of desired size (30-42 kb) were then precipitated from the sucrose solutions, dephosphorylated, and ligated into the *Bam*HI site of the SuperCos-1 vector.

The third genomic library was constructed using the CopyRight™ Cloning Kits (Lucigen, Middleton, WI) following the manufacturer' instructions. Chromosomal DNA from PKB1 strain was partially digested with *Sau*3AI and size-fractioned on a 0.8% low-melting-point agarose gel. The region containing DNA fragments of 35 to 45 kb was excised and recovered from the gel by agarase digestion. After precipitation with isopropanol, the purified genomic DNA fragments were ligated with *Bam*HI-digested, dephosphorylated pSMART-FOS fosmid vector at 16°C for 16 h. The ligation reaction (10 µl) contained 70 ng insert DNA, 70 ng fosmid vector, 1 mM ATP, and 1 unit of BAC Clone™ Ligase in the 1x BAC Clone™ Ligation buffer, from which 4 µl of the resulting ligation mixture was packaged in vitro with Gigapack III XL packaging extract (Stratagene) followed by transfection into the *E. cloni*® Replicator™ host cells.

2.4.8 Creation of an Apra^{R}Cm^R*oriT* disruption cassette

The *P. polymyxa fusA*, *ymcC* and *ktp* mutants were prepared using a modified version of the recently described Redirect© PCR targeting system (Gust *et al.*, 2003). None of the antibiotic resistance cassettes included as part of the original Redirect© system (Plant Biosciences, Norwich, UK) were suitable for use with *P. polymyxa* directly, and so a new plasmid construct, pJL9, carrying a hybrid gene disruption cassette Apra^{R}Cm^R*oriT*, was prepared. Briefly, the Apra^R*oriT* [*aac(3)IV* and *oriT*$_{\mathrm{RK2}}$] cassette from pIJ773 was cloned into pIJ2925 as a *Hin*dIII-*Eco*RI fragment to give the plasmid pJL7. The chloramphenicol resistant gene *cat* was removed from pC194 as a *Sau*3A-*Hpa*II fragment and inserted between the *Bam*HI and *Acc*I sites of pIJ2925 to generate pJL1. The *cat*-containing *Pst*I-*Eco*RI fragment of pJL1 was then cloned into pBluescript SK+, from where the *cat* gene was excised as a *Sma*I fragment and inserted into a unique *Nae*I site located between the *oriT* and *aac(3)IV* (Apra^R) genes in the plasmid pJL7, to produce the plasmid pJL9. The new hybrid resistance cassette (Apra^{R}Cm^R*oriT*) imparted both apramycin resistance for selection in *E. coli* and chloramphenicol resistance for selection in *P. polymyxa*.

2.4.9 PCR-targeted gene disruption

The Apra^{R}Cm^R*oriT* disruption cassette was amplified by PCR using a gel-purified 2.5-kb *Bgl*II fragment from pJL9 as template. Each primer consists of a 39-nt targeting sequence at the 5' end, identical to regions flanking the target gene, and a 19- or 20-nt priming sequence at the 3' end, derived from the disruption cassette. The reaction (50 µl) was set up using 50 ng gel-purified DNA fragment, 50 pmol of each primer, 0.5 µl Taq DNA polymerase, and 1 µl of 1:50 diluted Pfu DNA polymerase (Fermentation Service

Unit, Department of Biological Sciences, University of Alberta) in 1x Taq buffer. Amplification was conducted under the following PCR conditions: 94°C for 2 min; 10 cycles of 94°C for 45 s, 50°C for 45 s, and 68°C for 2 min; 15 cycles of 94°C for 45 s, 55°C for 45 s, and 68°C for 2 min with the extension time increased by 2 s per cycle; and a final elongation step of 68°C for 5 min. The linear PCR product was introduced by electroporation into *E. coli* BW25113 carrying both pIJ790 and the cosmid carrying the target gene, wherein the λ Red functions encoded by pIJ790 promoted homologous recombination between the target gene and the PCR-amplified disruption cassette. To prepare *ymcC* and *ktp* mutants, the λ Red recombination plasmid pKD46 was used as a substitute for pIJ790, along with the use of fosmid 4G9, while all other conditions remained the same. The resulting mutagenized cosmid/fosmid, in which the target gene was replaced by the disruption cassette, was then transformed into *E. coli* ET12567/pUZ8002, and from there introduced into wild-type *P. polymyxa* PKB1 strain via intergeneric conjugation.

2.5 Protein analysis

2.5.1 Overproduction and purification of the internal adenylation domain FusA-A6

A DNA fragment encoding the sixth A domain (FusA-A6) was amplified from fosmid 6D11 using the Expand High Fidelity PCR system (Roche) with primers JRL37 and JRL38 (Section 2.4.2). The PCR product was first cloned into pCR®2.1-TOPO® (Invitrogen), and then excised as an *Nde*I-*Bam*HI fragment, which was ligated into a similarly digested vector pET-19b (Novagen, Darmstadt, Germany). The recombinant expression plasmid was transformed into *E. coli* BL21(DE3) and one transformant was grown in 2.5 ml LB medium containing 100 μg/ml ampicillin for 16 h at 37°C. The whole 16 h culture was added into 50 ml LB medium supplemented with 100 μg/ml ampicillin in a 250 ml flask, which was then incubated at 37°C with shaking at 300 rpm until the OD_{600} reached ~0.6. When induced with 1 mM IPTG, the culture yielded an N-terminally His-tagged FusA-A6 protein after an additional 4 h of growth at 22°C. Purification of soluble His-tagged FusA-A6 protein was performed by Ni^{2+}-affinity chromatography (Qiagen) according to the manufacture's protocol. Briefly, the 50 ml culture was harvested by centrifugation at 4,400 x g for 20 min. The cell pellet was frozen in a dry ice-ethanol bath, thawed on ice, and resuspended in 5 ml lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, pH 8.0) supplemented with 1 mg/ml lysozyme. The cell suspension was incubated on ice for 30 min and then sonicated on ice using a 2.5 mm probe (Branson Sonifier 450) for 10 x 10 s on low setting. The cell lysate was centrifuged at 10,000 x g for 20 min at 4°C, and 4 ml of the supernatant was mixed with 1 ml 50% Ni-NTA slurry in a 25 ml flask, which was incubated at 4°C for an hour with shaking at 200 rpm. The resulting mixture was loaded into a 5 ml plastic syringe barrel plugged with cotton wool, and endogenous proteins that unspecifically bound to the Ni-NTA matrix were removed by washing the column three times with 4 ml wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol, pH 8.0). The purified His-tagged protein was then eluted with 4 x 0.5 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol, pH 8.0). The second and third eluates were pooled together and desalted into assay buffer (50 mM sodium phosphate, pH 8.0, 10 mM $MgCl_2$, 2 mM dithiothreitol, 1 mM EDTA) using a NAP™-5 column (Amersham Pharmacia Biotech, Uppsala, Sweden). The purity of the recombinant protein was checked by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis with Coomassie Brilliant Blue staining. Protein concentration was determined by using a calculated ε_{280} for purified FusA-A6 of 65,375 $M^{-1}cm^{-1}$ (Pace *et al.*, 1995).

2.5.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were examined by electrophoresis on a 10% SDS-polyacrylamide gel using a Mini-Protein® III Electrophoresis Cell (Bio-Rad). The resolving gel was prepared with 10% Acrylamide/Bis solution (37.5:1, Bio-Rad), 0.1% SDS, and 0.375 M Tris-HCl (pH 8.8), and each 10 ml gel solution was polymerized with 5 µl TEMED (N,N, N',N'-tetra-methylethylenediamine, Bio-Rad) and 50 µl of 10% ammonium persulfate (Sigma). The stacking gel is composed of 4% Acrylamide/Bis solution (37.5:1), 0.1% SDS, and 0.125 M Tris-HCl (pH 6.8), and each 5 ml gel solution was polymerized with 25 µl of 10% ammonium persulfate and 5 µl TEMED. Samples were prepared for electrophoresis by mixing with equal volumes of 2x loading buffer (0.225 M Tris-HCl, pH 6.8, 50% glycerol, 5% SDS, 0.05% bromophenol blue, 0.25 M dithiothreitol) and boiling for 5 min. PageRuler™ Prestained Protein Ladder Plus (Fermentas) was used as the molecular weight marker. The gel was run at room temperature in 1x running buffer (0.3% Tris base, 1.44% glycine, 0.1% SDS, pH 8.3) at 200 V for 45-60 min, and the protein bands were visualized by staining the gel with Coomassie Brilliant Blue as described by Sambrook *et al.* (1989).

2.5.3 ATP-PPi exchange assay

The amino acid specificity of purified recombinant FusA-A6 protein was determined using an ATP-PPi exchange assay, performed as previously described by Mootz and Marahiel (1997) with minor modifications. Reaction mixtures contained 50 mM sodium phosphate (pH 8.0), 2 mM $MgCl_2$, 2 mM dithiothreitol, 2 mM ATP, 1 mM amino acid substrate, 0.2 μCi tetrasodium [^{32}P] pyrophosphate (Perkin Elmer, Boston, MA) and 0.1 mM tetrasodium pyrophosphate. Reactions were initiated by the addition of recombinant FusA-A6 protein to a final concentration of 250 nM in a total volume of 100 μl. After incubation at 37°C for 15 min, reactions were stopped by adding 0.5 ml of termination mixture (1.2% [w/v] activated charcoal, 0.1 M tetrasodium pyrophosphate, and 0.5 M $HClO_4$). The charcoal was sedimented by microcentrifugation at 14,000 rpm for 5 min, washed once with 1 ml distilled water, and resuspended in 0.5 ml distilled water. The samples were then added to scintillation vials containing 4 ml scintillation fluid, and the radioactivity bound to the charcoal was determined using the Beckman LS 3801 liquid scintillation counter.[12]

65

---

[1] *A version of this chapter has been published. Li et al. 2007. Applied and Environmental Microbiology. 73: 3480-3489.*

[2] *A version of this chapter has been submitted for pulication. Li and Jensen 2007.*

Table 2.1 Bacterial and fungal strains used in this study.

Strain	Relevant description [a]	Source / reference
***E. coli* strains**		
BL21(DE3)	Host for expression of recombinant His-tagged FusA-A6 protein, *hsdS gal dcm* (DE3)	Novagen
BW25113	Recombinant host for PCR-targeting mutagenesis	Datsenko and Wanner (2000)
DH5α	Host for general cloning experiments and preparation of plasmid DNA	Gibco BRL
E. cloni® Replicator™	Host for the pSMART-FOS genomic library and preparation of recombinant fosmids	Lucigen
ER1447	Methylation-deficient host, *dam13*::Tn9 *dcm6 mcrB1 mcrA hsdR2*, Str^{R}, Cm^{R}	J. McCormick, Harvard University
ET12567	Methylation-deficient host, *dam13*::Tn9 *dcm6 hsdR*, Tet^{R}, Cm^{R}	MacNeil *et al.* (1992)
One Shot® TOP10	Host for TOPO® cloning reaction	Invitrogen
XL1-Blue	Host for general cloning experiments and preparation of plasmid DNA	Statagene
XL1-Blue MR	Host for the SuperCos-1 genomic libraries and preparation of recombinant cosmids	Stratagene

Table 2.1 (continued)

Strain	Relevant description [a]	Source / reference
***P. polymyxa* strains**		
PKB1	Wild-type strain producing fusaricidins	P. Kharbanda, Alberta Research Council
A1 /A5	*fusA* disruption mutants (single crossover), Apra^{R}, Cm^{R}	This study
A4 / A6	*fusA* disruption mutants (double crossover), Apra^{R}, Cm^{R}	This study
#3 / #12	*ymcC* disruption mutants (single crossover), Apra^{R}, Cm^{R}	This study
#8 / #11	*ymcC* disruption mutants (double crossover), Apra^{R}, Cm^{R}	This study
T1	*ktp* disruption mutant (double crossover), Apra^{R}, Cm^{R}	This study
Fungal strain		
L. maculans	Indicator organism for fusaricidin bioassay	P. Kharbanda, Alberta Research Council

[a] Apra^{R}, apramycin resistance; Cm^{R}, chloramphenicol resistance; Str^{R}, streptomycin and spectinomycin resistance; Tet^{R}, tetracycline resistance.

Table 2.2 Cloning vectors and recombinant plasmids used in this study.

Plasmid	Relevant description [a]	Source / reference
pBEST501	*E. coli* plasmid carrying a neomycin resistance gene cassette selectable in *Bacillus subtilis*. Amp^R, Neo^R	Itaya *et al.* (1989)
pBEST502	*E. coli* plasmid carrying the same neomycin resistance gene cassette as pBEST501, but flanked by different restriction sites, Amp^R, Neo^R	Itaya *et al.* (1989)
pBluescript SK+	*E. coli* phagemid cloning vector, Amp^R	Stratagene
pBM-SPECT	*E. coli* plasmid carrying a spectinomycin resistance gene cassette selectable in *Bacillus subtilis*, Amp^R, Spec^R	S.L. Wong, University of Calgary
pC194	*Staphylococcus* plasmid, also capable of replication in *P. polymyxa*. Cm^R	Horinouchi and Weisblum (1982)
pC194-*apra*	*E. coli-Paenibacillus* shuttle plasmid: pC194 ligated to *Hin*dIII-digested pIJ773	This study
pCR®2.1-TOPO®	*E. coli* phagemid vector for direct cloning of PCR-amplified fragments, Amp^R, Kan^R	Invitrogen
pET-19b	*E. coli* cloning vector for overexpression of N-terminal His-tagged recombinant proteins, Amp^R	Novagen
pIJ2925	pUC18 derived vector containing two *Bgl*II sites flanking the multiple cloning site, Amp^R	Janssen and Bibb. (1993)
pIJ773	pBluescript KS+ containing the Apra^R*oriT* [*aac(3)IV* and *oriT*$_{RK2}$] gene disruption cassette, Amp^R, Apra^R	Gust *et al.* (2003)

Table 2.2 (continued)

Plasmid	**Relevant description** ^{a}	**Source / reference**
pIJ790	λ Red expression plasmid, *gam bet exo araC rep101*^{ts}, Cm^{R}	Gust *et al*. (2003)
pJH101	*E. coli*-based integrational vector, Amp^{R}, Tet^{R}, Cm^{R}	Ferrari *et al*. (1983)
pJL1	pIJ2925 carrying the *cat* gene from pC194, Amp^{R}, Cm^{R}	This study
pJL5	pBluescript SK+ carrying the *cat*-containing *Pst*I-*Eco*RI fragment of pJL1, Amp^{R}, Cm^{R}	This study
pJL7	pIJ2925 carrying the Apra^{R}*oriT* gene disruption cassette from pIJ773, Amp^{R}, Apra^{R}	This study
pJL9	pIJ2925 carrying the Apra^{R}Cm^{R}*oriT* gene disruption cassette, Amp^{R}, Apra^{R}, Cm^{R}	This study
pJL-asn:*cat*	pUC119Ω carrying the *cat*-disrupted *fus-asn* gene fragment, Str^{R}, Cm^{R}	This study
pJL-asn:JH101	pJH101 carrying the *fus-asn* gene fragment, Amp^{R}, Tet^{R}, Cm^{R}	This study
pKD46	λ Red expression plasmid, *gam bet exo araB rep101*^{ts}, Amp^{R}	Datsenko and Wanner (2000)
pSETΩ	*E. coli* plasmid carrying *aadA*, the spectinomycin-streptomycin resistance gene selectable in *Streptomyces*, Str^{R}, Spec^{R}	O'Connor *et al*. (2002)

Table 2.2 (continued)

Plasmid	Relevant description [a]	Source / reference
pSKNeo-S	pBluescript II SK+ carrying a neomycin resistance gene cassette selectable in *Streptomyces*, Amp^R, Neo^R	Jensen's lab
pSMART-FOS	*E. coli* fosmid cloning vector for PKB1 genomic library preparation, single *cos* site, *parABC repE oriS oriV*, Cm^R	Lucigen
pUC118	*E. coli* phagemid cloning vector, Amp^R	Vieira and Messing (1987)
pUC119	*E. coli* phagemid cloning vector, Amp^R	Vieira and Messing (1987)
pUC119Ω	pUC119 derivative in which the *bla* gene (Amp^R) was replaced by the *aadA* gene from pSETΩ, Str^R, Spec^R	This study
pUZ8002	A non-transmissible plasmid encoding transfer function for mobilizing *oriT*-containing plasmid/cosmid/fosmid from *E. coli* to *P. polymyxa*, *tra*, RP4, Kan^R	Paget *et al*. (1999)
SuperCos-1	*E. coli* cosmid cloning vector for PKB1 genomic library preparation, dual *cos* sites, Amp^R, Kan^R	Stratagene

[a] Amp^R, ampicillin resistance; Apra^R, apramycin resistance; Cm^R, chloramphenicol resistance; Kan^R, kanamycin resistance; Neo^R, neomycin resistance; Str^R, streptomycin resistance; Spec^R, spectinomycin resistance; Tet^R, tetracycline resistance.

Table 2.3 Oligonucleotide PCR primers used in this study.

Primer	Sequence (5'-3') [a]	Use / function [b]
B1	CCG GCC GCT GCG GIT G(CT)(AT) (CG)IA C	Primer pair for amplifying peptide synthetase genes from the PKB1 chromosome
J6	ATG AGA ATT CTA GAG CTC IGA (AG)TG ICC ICC (AC)AG	
JRL1	TAC GGT GGG GCC TAT CTG ACT AT	Primer pair for PCR verification of *fus-asn* gene disruption
JRL3	TCT GCT ATA CTG GGA TGC GCT AAA	
JRL15-RD	CCT GAT GGC AAT ATT GAA TAT TTG GGG CGG ATC GAC CAT ATT CCG GGG ATC CGT CGA CC	Primer pair for PCR-targeting mutagenesis of *orf7* (or *fusA*)
JRL16-RD	TTT GAC TAG ACG GAC CAC GTT CCG GTG TTC AAC CAT AAC TGT AGG CTG GAG CTG CTT C	
JRL18'	GCG CAT CCA GCA AAT ATA CAG CCA CGT T	Primer pair for amplification of the TOPO-19 fragment
JRL19'	AGC CTG TCG TTT GAT GCA TCG GTT TCG	
JRL14	TCG GCC GGA TTT GAC GTC TGA GAA	Primers for PCR verification of *orf7*/*fusA* disruption
JRL30	TCC GTC TCC TCC ACC AAA GCC TCT	

Table 2.3 (continued)

Primer	Sequence (5'-3')	Use/function
JRL31-RD	CTC TTT TTC ATA AGA ACG GAT GGA GAG AAT ACT CTA ATG ATT CCG GGG ATC CGT CGA CC	Primer pair for PCR-targeting mutagenesis of *ktp*
JRL32-RD	GCT AAT CAG CAC GGG TAC ATC CTT TTT ATA GAT ACA TTA TGT AGG CTG GAG CTG CTT C	
JRL35	GCC GCC AAG CCA GTA GCA GTC GT	Primers for PCR verification of *ktp* disruption
JRL36	CGA ACC TTG GGC GCG CTG AAC	
JRL40	GGT GGA CTG GGC ATC GGC ATC G	
JRL37	TAA GGA TCC CCA AAT CTG CGC TAG TTC TAC	Primer pair for amplification of a DNA fragment corresponding to FusA-A6
JRL38	CTG CAT ATG CGT ATT GAT GAG CTG GAG TTG A	
JRL42-RD	AAA TAC CAA TTT CTA ATT TGA AAG GAA TCA TCT ATT ATG ATT CCG GGG ATC CGT CGA CC	Primer pair for PCR-targeting mutagenesis of *ymcC*
JRL43-RD	ACC AGC CAG ATA TCA TCT AAG TGT AAG TCT TAA CCT TTA TGT AGG CTG GAG CTG CTT C	
JRL44	GGG CAT GAA AGG GGG ATA GTT GAG GTA	Primers for PCR verification of *ymcC* disruption
JRL45	GGT AGC GAG GCA AAA ACA AGG TGA CTG	

Table 2.3 (continued)

Primer	**Sequence (5'-3')**	**Use/function**
SEJ18	CCG GCG GTG TGC TGC TGG TC	Primer pair for amplification of an *aac(3)IV*–specific fragment
SEJ19	CGG C[illegible] CGC ATT CTT CGC ATC C	

[a] Alternative nucleotides at a particular position are shown in parentheses; engineered restriction sites are underlined.

[b] The usage of each PCR primer is described in details in Sections 2 & 3.

3. Results and discussion

Strains of *P. polymyxa* are known to produce various peptide antibiotics presumably synthesized by nonribosomal peptide synthetases (NRPSs). In addition to fusaricidins, three larger peptide metabolites were also detected in PKB1 liquid culture (Beatty and Jensen, 2002). This suggests that there may be at least three separate peptide synthetase gene clusters on the chromosome in addition to the putative gene cluster encoding the peptide synthetase responsible for fusaricidin biosynthesis. To clone the fusaricidin biosynthetic gene cluster (*fus*), peptide synthetase gene fragments were amplified from PKB1 to serve as probes for screening genomic DNA libraries. The highly conserved core motifs located in the adenylation and thiolation domains of NRPSs were targeted for degenerate primer design. In addition, Stachelhaus *et al.* (1999) and Challis *et al.* (2000) previously identified amino acid residues within the adenylation domains of peptide synthetase modules that are involved in amino acid substrate recognition. Accordingly, the PCR primers used for probe preparation were designed for amplification of a fragment encompassing the specificity-conferring codons of an A domain. From the resulting PCR products, two possible *fus*-specific NRPS gene fragments were selected as probes based on the predicted substrate specificity of the derived A domains. A total of three PKB1 genomic libraries were prepared and screened, and finally, the entire *fus* gene cluster was cloned and characterized. In addition, a modified PCR-targeting mutagenesis protocol was developed to create targeted mutations on the chromosome of *P. polymyxa* PKB1. This allowed verification of the involvement of *fusA*, the single peptide synthetase gene identified in the *fus* cluster, in the nonribosomal synthesis of fusaricidins, and determination of possible boundaries of the *fus* gene cluster. Taken together, this study provides the basis for genetic manipulation of fusaricidin production in *P. polymyxa* PKB1, and furthermore, for construction of novel antibiotics by combinatorial biosynthesis.

3.1 Analysis of the peptide synthetase gene fragments amplified from *P. polymyxa* PKB1

In order to clone the fusaricidin biosynthetic genes, degenerate oligonucleotide primers corresponding to two conserved core motifs, A2 and T, found within peptide synthetases, were designed by Dr. P. H. Beatty and colleagues at Alberta Research Council (Figure 3.1.1). *P. polymyxa* PKB1 genomic DNA was used as template for PCR amplification of putative peptide synthetase gene fragments. The colleagues at Alberta Research Council also kindly cloned and sequenced all PCR products. We analyzed fifteen different PCR-amplified fragments of the expected size, approximately 1.65 kb, and all of them gave terminal DNA sequence information indicating that they were derived from peptide synthetase genes. Eight of the fragments gave sufficient sequence information to allow analysis of the internal core motifs, A3 to A10, in their deduced amino acid sequences. According to the selectivity-conferring codes identified within the adenylation (A) domains of NRPSs (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999), substrate specificities of the A domains derived from these eight PCR-amplified fragments were predicted (Table 3.1). On this basis, DNA fragments B1J6-2, B1J6-3, B1J6-5, B1J6-14, and B1J6-19 apparently encode peptide synthetase modules that specifically recognize glutamic acid or ornithine. Neither of these amino acids has been identified as a component of fusaricidin, suggesting that these fragments were not derived from the peptide synthetase gene cluster for fusaricidin. Since *P. polymyxa* strains are known to produce a variety of peptide metabolites in addition to fusaricidin, it was not unexpected to isolate NRPS genes for other peptide products. The DNA fragment B1J6-10 showed high similarity to the Asn-activating A domain from tyrocidine synthetase TycC (GenBank accession no. AAC45930), and appeared to be the most suitable probe for subsequent screening of a PKB1 genomic DNA library since asparagine was considered a unique residue of fusaricidins and it has not been found in any other peptide metabolites isolated from *P. polymyxa* so far. The deduced amino acid sequences of the other two PCR-amplified fragments, B1J6-7 and B1J6-17, displayed high levels of homology with adenylation domains activating threonine. Since all fusaricidin analogs have invariant threonine (or *allo*-threonine) residues at the first and fourth positions of the peptide moiety (Figure 1.3.1), the fragment B1J6-17, the more completely defined of the two Thr-specific probes, was chosen as a second probe for library screening. This fragment exhibits about 80% identity to the Thr-activating module of the fengycin synthetases from *Bacillus subtilis* (GenBank accession no. CAA09819).

3.2 Screening of the first SuperCos-1 genomic library of *P. polymyxa* PKB1

SuperCos-1 cosmid vector is an efficient tool in genomic sequence analysis because of its capacity to accommodate large DNA fragments (30-42 kb). The first SuperCos-1 genomic library of PKB1 was constructed by partial *Sau*3AI digestion of genomic DNA, dephosphorylation of the insert fragments, and ligation into the unique *Bam*HI site of the cosmid vector. The resulting ligation mixture was packaged *in vitro*, and then transfected into *E. coli* XL1-Blue MR cells for amplification. The insert DNA in recombinant cosmids can be excised using the *Not*I or *Eco*RI restriction sites flanking the *Bam*HI cloning site (*Eco*RI / *Not*I / *Bam*HI / *Not*I / *Eco*RI). Due to the potential presence of multiple NPRS gene clusters on the chromosome of *P. polymyxa*, the cosmid library was screened with two different PCR-amplified peptide synthetase gene fragments as probes, in an effort to isolate cosmids containing NRPS genes from the fusaricidin biosynthetic gene cluster.

3.2.1 The first screening of PKB1 genomic library, with the B1J6-10 fragment

In a preliminary analysis of the SuperCos-1 genomic library, ten randomly-selected recombinant cosmids were digested with *Not*I restriction enzyme to assess their insert sizes. It was shown that all of these cosmids contained DNA inserts of above 23 kb, suggesting that most clones in the cosmid library may have large size inserts. A total of ca. 2,000 colonies from the genomic library were then screened at moderate stringency, using B1J6-10 (the Asn-specific fragment) as a probe. Positive clones hybridizing with the probe were subjected to colony hybridization with B1J6-17 (the Thr-specific fragment) as a probe. This was done to identify cosmids carrying peptide synthetase genes which encode both Asn-activating and Thr-activating modules, as expected for fusaricidin biosynthesis. Twenty-six positive cosmids obtained from the second round of hybridization were digested with *Eco*RI, electrophoresed on an agarose gel, transferred onto a nylon membrane and hybridized with the fragments B1J6-10 and B1J6-17, respectively. This Southern analysis was used to distinguish insert size and restriction pattern. Only one cosmid, designated Col-8 (with an insert of ~39 kb), hybridized to both probes and was selected for further study. The cosmid Col-8, together with genomic DNA from the wild-type PKB1 strain, were digested with various restriction enzymes and analyzed by Southern hybridization (Figure 3.2.1). When the membrane blot was probed using the Asn-specific fragment B1J6-10, a 3.2-kb *Eco*RI fragment from Col-8 strongly hybridized to the probe, whereas the corresponding fragment was not seen in the *Eco*RI digestion of genomic DNA. Instead, the genomic digest had a smaller *Eco*RI fragment of 2.2 kb that hybridized to the same probe, B1J6-10. This discrepancy suggests that the insert DNA of Col-8 may have rearranged or contain non-contiguous chromosomal fragments.

The two hybridizing *Eco*RI fragments from either Col-8 or the PKB1 chromosome were subcloned and fully sequenced. The 2.2-kb *Eco*RI fragment was obtained by subcloning a mixture of fragments between 1.9 kb and 2.3 kb from the *Eco*RI-digested genomic DNA and subsequent screening the plasmid library using the fragment B1J6-10. By detailed analysis of the deduced amino acid sequence, an incomplete Asn-activating module typical of NRPSs was identified in the 3' portion of the 3.2-kb *Eco*RI fragment from Col-8, whereas its 5' portion shows homology with phosphotransferase system (PTS) glucose-specific components. The 2.2-kb genomic *Eco*RI fragment also encodes a partial peptide synthetase module, consisting of a partial condensation domain, a putative Asn-activating adenylation domain, a full thiolation domain, and a partial epimerization domain. Alignments of the nucleotide sequences of these two *Eco*RI fragments with the PCR-amplified fragment B1J6-10 revealed that B1J6-10 shares the same DNA sequence with the 2.2-kb *Eco*RI fragment, indicating direct amplification of B1J6-10 from this locus on the chromosome, and that the two *Eco*RI fragments have the same sequence in the 3' portion, but not in the 5' portion. It was apparent that the 3.2-kb *Eco*RI fragment may contain at least two discontinuous genomic DNA fragments. By end sequencing of Col-8 insert DNA, the 3.2-kb *Eco*RI fragment was found to reside at one end of the insert, with the PTS gene adjacent to the SuperCos-1 vector (Section 3.2.3). This result led to the hope that sequence discontinuity was restricted to the end of the cosmid insert.

Among the known peptide metabolites produced by *P. polymyxa*, fusaricidins are the only peptide products containing D-Asn as a component. The high level of similarity between the 2.2-kb *Eco*RI fragment and the D-Asn-incorporating module from mycosubtilin synthetase MycB (GenBank accession no. AAF08796) suggests that this fragment is likely to be part of the *fus* gene cluster and thus was designated as *fus-asn*. Efforts were taken to investigate the role of *fus-asn* in fusaricidin biosynthesis by disruption of the chromosomal copy of *fus-asn* in *P. polymyxa* PKB1 (Section 3.5).

3.2.2 The second screening of PKB1 genomic library, with the B1J6-17 fragment

Fusaricidin represents a mixture of several lipopeptides, in which the residues at the second, third and fifth positions in the peptide portion are highly variable. Amino acid substitutions at defined positions were also reported for many other nonribosomally synthesized peptides. It was previously demonstrated that the amino acid composition of the peptide products can be affected by changing the relative concentrations of certain amino acids in the growth media (Ruttenberg and Mach, 1966). The diversity in the peptide sequence can also be related to the relatively broad substrate specificities of the respective A domains of NRPSs in comparison with those of aminoacyl-tRNA synthetases. For example, in the tyrocidine synthetase TycC, the A domain in the third module was found to activate and incorporate L-Trp in position seven of the resulting peptide to give tyrocidine variant D, while the same position in tyrocidine A, B and C is replaced by an L-Tyr residue, indicating that L-Tyr can be recognized and activated by an A domain that displayed substrate specificity primarily for L-Trp (Mootz and Marahiel, 1997). That is, the predicted substrate specificity based on the specificity-conferring code of an A domain may not necessarily be consistent with the actual occurrence of amino acids in the corresponding position of the peptide product. Accordingly, it was hypothesized that the A domain responsible for incorporation of the asparagine or glutamine residue in the fifth position of fusaricidin peptides (Figure 1.3.1), may activate both amino acid residues, but the specificity-conferring code of this A domain could be similar to that of a Gln-activating domain rather than an Asn-activating one. If this is the case, the *fus-asn* fragment, encoding a partial Asn-activating module, may not be part of the fusaricidin biosynthetic gene cluster, but rather part of a NRPS gene cluster for some as yet uncharacterized Asn-containing peptide metabolites from *P. polymyxa*. Under those circumstances, using the Asn-specific fragment B1J6-10 as the only probe for screening the genomic library could exclude identification of cosmids carrying the fusaricidin synthetase genes. Therefore, the remaining PCR-amplified NRPS gene fragments were examined again, and the fragment B1J6-17 with high similarity to Thr-activating modules was selected as the only suitable probe for a second screening of the cosmid library. Although threonine residues are present in several of the other peptide metabolites produced by *P. polymyxa,* as well as in fusaricidins, none of the other PCR-produced probes exhibit substrate specificities that are consistent with the primary structure of fusaricidins.

As expected, screening the genomic library with the B1J6-17 probe yielded many positive clones. By restriction analysis and Southern hybridization, five cosmids Col-9, Col-10, Col-15, Col-19 and Col-20 were isolated, which hybridized with the probe but showed restriction patterns different from each other and from that of Col-8, suggesting that they were distinct cosmid clones. In addition, Southern analyses revealed that cosmids Col-8, Col-9, Col-10, Col-15, and Col-20 had similar restriction fragments that hybridized with the B1J6-17 probe, while distinct hybridizing bands were detected in the other cosmid Col-19 (Figure 3.2.2). These findings suggest that the six cosmids could be divided into two main groups. Five cosmids Col-8, Col-9, Col-10, Col-15, and Col-20, probably contain overlapping DNA inserts and thus comprise group 1. The single cosmid Col-19, making up group 2, apparently contains genes from a separate locus on the chromosome, indicating that two distinct peptide synthetase gene clusters were identified in the PKB1 strain. A representative member of the first group of clones, Col-8, and the only member of the second group of clone, Col-19, were selected for further study, and the complete sequences of their DNA inserts were determined by a commercial DNA sequencing agency (SeqWright).

3.2.3 Sequence analysis of the insert DNA of cosmid Col-8

The 39-kb insert of Col-8 (Appendix A) covers four complete and four partial open reading frames (ORFs), all in the same transcriptional direction (Figure 3.2.3A). Sequence analysis revealed that the 3.2-kb *Eco*RI fragment previously subcloned from Col-8 resides at the extreme 5' end of the insert DNA, and overlaps with *orf1* and *orf2*, encoding a partial phosphotransferase system component and an incomplete peptide synthetase module, respectively (Figure 3.2.3, Table 3.2).

Downstream of *orf2* were four ORFs, *orf5*, *orf6*, *orf7*, and *orf10*, homologous to known NRPS genes (Figure 3.2.3, Table 3.2). By comparison with the specificity codes extracted from other peptide synthetases, it was predicted that the amino acid substrates, most likely to be recognized by the A domains within the deduced amino acid sequences of these four ORFs, are threonine, isoleucine, and ornithine (Table 3.3). Since neither isoleucine nor ornithine is present in fusaricidin, the peptide synthetase gene fragments identified in Col-8 are unlikely to be part of the fusaricidin biosynthetic gene cluster. However, none of the peptide metabolites described from *P. polymyxa* isolates contain ornithine (2,5-diaminovaleric acid). Instead, the closely related compound DAB (2,4-diaminobutyric acid) is found in several *P. polymyxa* peptide products, but not in fusaricidin. Therefore, the modules identified as incorporating ornithine may actually represent DAB-activating modules. By analyzing the deduced amino acid sequences, I identified a putative Ile-activating module followed by two putative Orn-activating modules within *orf6*, and a putative Thr-activating module flanked on either side by a putative Orn-activating module within *orf10*. Obviously, the amino acid sequence, Ile-DAB-DAB or DAB-Thr-DAB, is inconsistent with the structure of fusaricidin. However, polymyxin, a well-known peptide antibiotic from *P. polymyxa*, is the only one of the DAB-containing products of this species to include amino acid sequences of both Ile-DAB-DAB and DAB-Thr-DAB (Figure 3.2.4). In addition, sequence analysis of *orf7* revealed the presence of a putative Thr-activating module including a C-terminal TE domain, which matches the occurrence of L-Thr at the C-terminal position of the polymyxin peptide. Based on these findings, it was postulated that Col-8 and the other cosmids of the first group of clones may represent the polymyxin biosynthetic gene cluster, but additional experimentation is clearly required before that conclusion can be firmly established.

The predicted gene products of *orf8* and *orf9*, located between *orf7* and *orf10*, are homologs of ABC transporter proteins (Table 3.2). ABC-type transporters flanking peptide synthetases are believed to be required for extracellular export of the peptide product (Gaisser and Hughes, 1997). Hence, it was proposed that *orf8* and *orf9* may be involved in the secretion of polymyxins.

3.2.4 Sequence analysis of cosmid Col-19

Col-19, the single cosmid of the second group of clones that hybridized to the B1J6-17 probe, contains an insert DNA of 36.4 kb (Appendix B), and covers 11 complete and 3 partial open reading frames. The organization and proposed functions of the ORFs identified within Col-19 are shown in Figure 3.2.5 and Table 3.4, respectively.

3.2.4.1 Identification of a putative fusaricidin synthetase gene, *orf7*

Sequence analysis of Col-19 revealed the presence of an incomplete 9,750-bp open reading frame, *orf7*, encoding a peptide synthetase presumably involved in fusaricidin biosynthesis. This partial ORF starts with an ATG codon at nt 6508 and is preceded by a putative ribosome-binding site (AGGAG) located 8 bp upstream (Appendix B). Within the deduced amino acid sequence of *orf7* (designated ORF7), three modules typical of peptide synthetases were identified, although the third module is truncated at the C-terminus. As shown in Figure 3.2.6, a total of nine catalytic domains can be distinguished: three in the first module, four in the second, and two in the third (the last domain is incomplete). The predicted substrate specificities of the A domains in the first and second modules were consistent with the first two amino acids, Thr and Val, in the peptide moiety of fusaricidin (Table 3.3). In addition, the presence of an epimerization (E) domain at the C-terminal end of the second module indicated that the activated amino acid valine is likely to be converted into the D-configuration and accordingly, D-Val will be incorporated into the peptide product at position two. Furthermore, the N-terminal condensation (C) domain in the first module exhibited 58% similarity to the first C domain of the surfactin synthetase SrfA-A, which is believed to act as an acceptor for the fatty acid moiety of the lipopeptide, surfactin (Cosmina *et al.*, 1993) (see Section 3.6.1.1 for detailed discussion). Taken together, these findings suggest that the peptide synthetase encoded by *orf7* is responsible for incorporation of L-Thr in the first position, and D-Val in the second position of the peptide product, and the L-Thr residue is associated with a lipid moiety. All of these features are in agreement with the primary structure of fusaricidin, but given the large number of peptide metabolites known to be produced by *P. polymyxa* and the possibility of other as yet undescribed peptides, independent confirmation of the involvement of *orf7* in fusaricidin production was carried out (Section 3.5.4).

3.2.4.2 Cloning of a continuous chromosomal fragment downstream of *orf7*

Disruption of *orf7* completely abolished the antifungal activity of strain PKB1 (Section 3.5.4), indicating that it is part of the fusaricidin biosynthetic gene cluster (*fus*) and encodes the first three modules of fusaricidin synthetase. However, sequence analysis of Col-19 indicated that *orf7* is truncated at the 3' end, preventing complete analysis of the *fus* gene cluster. For sequencing the continuous chromosomal region downstream from *orf7*, long-distance PCRs were performed (Figure 3.2.6). The primer JRL19' was designed based on the 3' end sequence of *orf7*. The pairing primer JRL18' was derived from the *fus-asn* fragment, based on the assumption that the partial Asn-activating module encoded by *fus-asn* would be responsible for the incorporation of D-Asn in the fifth position of fusaricidin, and thus *fus-asn* would be located downstream of *orf7* on the chromosome. A PCR product of about 4.5 kb, designated TOPO-19, was obtained using PKB1 genomic DNA as the template, and then cloned into pCR®2.1-TOPO® vector for full sequencing. The deduced amino acid sequence of TOPO-19 showed high similarity to Thr-activating modules associated with an epimerization domain (Figure 3.2.6). This result suggests that TOPO-19 probably encodes the fourth module of fusaricidin synthetase, incorporating D-*allo*-Thr in position four of the peptide product, and so Col-19, like Col-8, appears to contain a noncontiguous DNA fragment not representative of the genome sequence.

Surprisingly however, it was found that TOPO-19 was actually amplified from two JRL19' primers, while the other primer JRL18' was not detected at either end of the PCR product. Despite repeated attempts varying conditions of PCR reactions, such as the concentrations of primers, enzyme and $MgCl_2$, the annealing temperature, the extension time and the number of cycles, it was never possible to amplify a DNA fragment that would link *orf7* to the *fus-asn* fragment. This observation implies that long-distance PCR reactions using primers JRL18' and JRL19' were unsuccessful, or that the *fus-asn* fragment is not located nearby the *orf7* region due to the unusual gene arrangement in the fusaricidin biosynthetic gene cluster. It also could be that the *fus-asn* fragment is not involved in fusaricidin biosynthesis as previously assumed (Section 3.2.2); instead, it may be part of a NRPS gene cluster synthesizing peptide metabolites that contain D-Asn residues, but have not been identified from *P. polymyxa* so far.

3.3 Preparation and screening of a second SuperCos-1 genomic library of *P. polymyxa* PKB1

The presence of discontinuous genomic DNA fragments within cosmids Col-8 and Col-19 indicated that rearrangement of the insert DNA had occurred in the first genomic library. Alternatively, small non-contiguous DNA fragments may have been randomly incorporated into cosmids during library construction, since the partially digested chromosomal DNA was not size-fractionated before ligation into the SuperCos-1 vector. Furthermore, even after two additional rounds of screening the first genomic library, Col-8 was the only positive clone that hybridized to both B1J6-10 and B1J6-17 probes, suggesting that genomic sequences of the PKB1 strain were not completely represented in this library, and thus it was not suitable for further investigation. Therefore, a second SuperCos-1 library of PKB1 was prepared with size-selected genomic DNA fragments.

The *Sau*3AI partially digested genomic DNA was fractionated by sucrose gradient centrifugation (Section 2.4.7). Fractions of the sucrose gradient were examined by agarose gel electrophoresis to determine which fractions contain DNA fragments between 30 kb and 42 kb. The DNA fragments of desired size were precipitated from the sucrose solution, dephosphorylated, and ligated into the *Bam*HI site of SuperCos-1 vector. Based on the assumption that the *fus-asn* fragment would still be part of the fusaricidin synthetase genes, it was used as a probe to screen approximately 3,000 colonies from the second genomic library. Twenty-four cosmids were obtained in the first screening of the library. These cosmids could be divided into four groups according to their restriction patterns; however, one group of clones shared the same restriction pattern with Col-8. Southern analysis showed that these four groups of clones gave four distinct *Eco*RI fragments hybridizing with the *fus-asn* fragment. Of the four hybridizing fragments that were subsequently cloned and fully sequenced, one *Eco*RI fragment had the same nucleotide sequence as the 3.2 kb *Eco*RI fragment from Col-8, while the other three fragments all consisted of truncated peptide synthetase genes fused to fragments of other unrelated genes, indicating that DNA rearrangement of cosmid inserts also occurred in the second genomic library, despite careful attention to size selection of insert DNA fragments during library construction. Therefore, it was proposed that the instability of peptide synthetase genes occurring in both cosmid libraries may be associated with the potential disadvantages of SuperCos-1 vector, and so a single-copy cloning vector was used for constructing a third genomic library of *P. polymyxa* PKB1.

3.4 Preparation and screening of a fosmid genomic library of *P. polymyxa* PKB1

SuperCos-1 vector has several advantageous features for generating genomic libraries for sequence analysis. It has a pUC origin of replication and is maintained at high copy number, giving rise to high cosmid yield. It also contains two bacteriophage promoters flanking the multiple cloning site, allowing for synthesis of insert end-specific

RNA probes for chromosomal walking. However, the high copy number of recombinant clones and vector-driven transcription into the insert DNA may cause difficulties in cloning of certain DNA sequences, such as toxic genes, large inserts, and AT-rich or highly repetitive DNA (Kieleczawa, 2005). To circumvent these apparent instability problems, a third genomic library of PKB1 was constructed using pSMART-FOS, a single-copy fosmid vector designed for stable maintenance of DNA inserts. Once genomic library construction was complete, addition of L-arabinose in the growth medium would indirectly activate an inducible medium-copy replication origin located in the vector, increasing fosmid yield up to 50 copies per cell. Another advantageous feature of this vector is the lack of transcription into or out of the insert. The presence of transcriptional terminators on both sides of the cloning site greatly stabilizes recombinant fosmids, and reduces the cloning bias against unstable targets.

To prepare a fosmid library, *P. polymyxa* PKB1 genomic DNA was partially digested with *Sau*3AI, and DNA fragments of 35 to 45 kb were selected and cloned into the *Bam*HI site of pSMART-FOS vector. A 1-kb *Eco*RI fragment obtained from the fragment TOPO-19, designated P-tyr, was used as a probe to screen 550 colonies from the fosmid library (Figure 3.2.6). Twenty-one hybridizing fosmids isolated in the first screening were digested with *Hin*dIII and used in Southern analysis. A 3.4-kb *Hin*dIII fragment hybridizing with the P-tyr probe was detected in four fosmids, 1C2, 3A9, 4G9, and 6D11, as well as in PKB1 genomic DNA, suggesting that these four fosmids probably contain overlapping inserts (Figure 3.4.1). However, 1C2 and 3A9 were excluded from further study as both fosmid inserts appeared to have rearranged during DNA amplification with addition of L-arabinose. Southern analyses showed no evidence of rearrangement of the insert DNA in the other two fosmids 4G9 and 6D11 (Figure 3.4.2), and thus these fosmids were sequenced by a commercial DNA sequencing service (SeqWright). In-house primer walking sequencing was also performed to fill in the gaps between DNA sequence contigs obtained from Seqwright. Finally, sequence analysis of both fosmid inserts yielded approximately 48 kb of contiguous DNA sequence information, which harbors the entire fusaricidin biosynthetic gene cluster (*fus*). However, before the fosmid sequences became available, a partial peptide synthetase gene, *orf7*, was identified on comid Col-19, which appeared to encode part of the fusaricidin synthetase (Section 3.2.4). Mutagenesis study was carried out to confirm the involvement of *orf7* in fusaricidin production (Section 3.5.4), and subsequent analysis of the *fus* gene cluster assembled from two fosmid inserts revealed that *orf7* is part of the fusaricidin synthetase gene *fusA* (Section 3.6.1).

3.5 Disruption of putative fusaricidin synthetase genes

In order to assess the importance of the peptide synthetase genes, *fus-asn* cloned from the PKB1 chromosome and *orf7* identified on cosmid Col-19, for fusaricidin biosynthesis, gene disruption experiments were performed. Since the genetic properties of *P. polymyxa* in general and the PKB1 strain in particular have not been fully characterized, attempts were made to identify antibiotic resistance markers that are capable of providing a clear selection at a single copy on the chromosome, and to develop efficient methods for introducing mutant alleles of the target genes into the chromosome of *P. polymyxa*.

3.5.1 Identification of selectable markers for use in *P. polymyxa* PKB1

A series of antibiotic resistance genes, functional in either *Bacillus* or *Streptomyces*, were tested for suitability of use in *P. polymyxa*. The plasmid pC194 was originally isolated from *Staphylococcus aureus*, and carries the chloramphenicol resistance gene *cat*. It has been proved to be capable of replication in *P. polymyxa* and to confer chloramphenicol resistance (Cm^R) on the cells (Bezzate *et al.*, 2000). Therefore, *E. coli* plasmids, carrying resistance markers to be tested, were fused to pC194 at the unique *Hin*dIII site to give *E. coli-Paenibacillus* shuttle vectors, which were then separately introduced into *P. polymyxa* PKB1 by electroporation. The susceptibility of PKB1 transformants harboring the resistance genes to the corresponding antibiotics at concentrations commonly used to inhibit *Bacillus* spp. was tested. At the same time, the minimum inhibitory concentrations of various antibiotics on the wild-type PKB1 strain were determined. Finally, in addition to the chloramphenicol resistance gene *cat* from pC194, the apramycin resistance gene *aac(3)IV* carried within the Apra^R*oriT* disruption cassette from pIJ773 (Section 3.5.4), and the streptomycin resistance gene *aadA* from the plasmid pSETΩ, also appeared to be useful antibiotic markers in *P. polymyxa*. None of the other antibiotic resistance markers that I have tested, such as the spectinomycin resistance marker from pBM-SPECT, the neomycin resistance genes from pBEST501, pBEST502, and pSKNeo-S, were suitable for use in *P. polymyxa* because they did not confer useful levels of resistance.

3.5.2 Transformation of *P. polymyxa* PKB1 with an integrational vector carrying the *fus-asn* gene fragment

An *E. coli*-based integrational vector, pJH101, was used to disrupt the *fus-asn* fragment in the chromosome of PKB1 *via* a single homologous recombination. The plasmid pJH101 contains the *cat* gene from pC194 and thus confers chloramphenicol resistance on *P. polymyxa*. Since it cannot replicate in *P. polymyxa*, chloramphenicol resistant transformants can only occur when the plasmid is integrated into the chromosome, which depends on the presence of a DNA fragment in the vector homologous to the target gene in the chromosome. The 2.2-kb *Eco*RI fragment *fus-asn* was originally obtained from the PKB1 chromosome, and encodes a partial peptide synthetase module presumably incorporating D-Asn into the peptide product. This DNA fragment was subcloned into a unique *Eco*RI site of pJH101 to give the integrational construct pJL-asn:JH101, which was then passed through a non-methylating *E. coli* strain, ER1447 or ET12567, before transformation into PKB1 cells by electroporation. Chloramphenicol resistant transformants were isolated, apparently resulting from single crossovers between the cloned copy and the chromosomal copy of *fus-asn*.

Genomic DNA from the wild-type PKB1 strain and four putative *fus-asn* mutants was digested with *Kpn*I and was used in Southern analysis to determine if pJH101 had integrated into the chromosome. The plasmid pJH101 was linearized by digestion at the unique *Eco*RI site, and the linearized plasmid was used as a probe in the subsequent hybridization experiment. Since neither pJH101 nor *fus-asn* contains a *Kpn*I restriction site, genomic DNA from the mutants was expected to produce a single hybridizing band that should not be present in the wild type. However, no fragments of the expected size (>9.8 kb) were seen to hybridize with the pJH101 probe in the lanes containing DNA from either mutants or wild-type PKB1 (Figure 3.5.1), indicating that these Cm^R single-crossover mutants may have reverted to the wild type by excision of the integrated vector via a second single crossover, or that they may actually represent spontaneous chloramphenicol resistant mutants of *P. polymyxa* PKB1. Even though more Cm^R transformants with the plasmid pJL-asn:JH101 were examined by Southern hybridization, no integrated copy of pJH101 on the chromosome was detected. Since no stable mutants generated by a single homologous recombination could be obtained, construction of a stable mutation of *fus-asn* by double crossover was attempted.

3.5.3 Insertional mutation of *fus-asn* via double crossover

The strategy used for generation of PKB1 mutants in which *fus-asn* is inactivated by insertion of an antibiotic resistance gene is outlined in Figure 3.5.2. The ampicillin resistance gene (*bla*) in the plasmid pUC119 was replaced by the *aadA* gene from the plasmid pSETΩ, which had been shown to confer streptomycin resistance (Str^R) on *P. polymyxa*, to give a new delivery vector pUC119Ω. The chloramphenicol resistance gene, *cat*, from the plasmid pC194 was cloned into the vector pIJ2925 to give the plasmid pJL1, from where the *cat* gene was excised as a *Bgl*II fragment. The *cat*-containing *Bgl*II fragment was then inserted into a unique *Bgl*II site located in the middle of the *fus-asn* fragment that had been cloned in pUC119Ω, to produce the gene disruption construct pJL-asn:*cat*. After passing through a methylation-deficient *E. coli* strain, ER1447 or ET12567, the plasmid pJL-asn:*cat* was introduced into PKB1 by electroporation. Chloramphenicol resistant transformants were selected and then grown in non-selective media to isolate chloramphenicol resistant but streptomycin sensitive mutants (Cm^R Str^S), which presumably would result from replacement of the chromosomal copy of *fus-asn* by the *cat*-disrupted copy via two homologous recombinations, one on either side of the *cat* marker.

The insertion of the *cat* gene in *fus-asn* on the chromosome was first examined by colony PCR. The primers JRL1 and JRL3 were designed to amplify a 1.4-kb fragment from the wild-type copy of *fus-asn*, while the *cat*-disrupted copy of *fus-asn* was expected to produce a larger PCR product of 2.4 kb (Figure 3.5.2). As shown in Figure 3.5.3, however, seven Cm^R Str^S mutants all gave a single PCR fragment that is the same size as the fragment amplified from the wild-type strain, indicating the absence of the *cat*-disrupted copy of *fus-asn* in their chromosomes. Southern analysis was also used to test for gain of the chloramphenicol resistance marker in the chromosome. When the *cat*-containing *Bgl*II fragment removed from the plasmid pJL1 was used as a *cat*-specific probe, no hybridizing bands were observed in the lanes containing genomic DNA from either wild-type PKB1 or the mutants (Figure 3.5.4). Taken together, these findings suggest that the Cm^R Str^S colonies obtained were likely to be spontaneous chloramphenicol resistant mutants of PKB1.

The *fus-asn* fragment was originally isolated from *P. polymyxa* PKB1 genomic DNA and therefore there should be sites of homology between the gene disruption construct pJL-asn:*cat* and the chromosome of PKB1. It was hypothesized that the homologous regions flanking the *cat* marker in the mutation construct pJL-asn:*cat* may be too short for recombination to occur, thus preventing selection of chloramphenicol resistant mutants resulting from double crossovers between the wild-type copy of *fus-asn* and the *cat*-disrupted copy. In general, the frequency of homologous recombination is proportional to the lengths of the homologous fragments used and so the use of large

DNA fragments in the gene disruption construct is recommended. However, the frequency of transformation by electroporation is constrained by the size of the plasmid to be transferred. Therefore, a more efficient mutation delivery method was required for introduction of large disruption construct into *P. polymyxa*.

3.5.4 Preparation of fusaricidin-nonproducer mutants using a modified Redirect© technology

Intergeneric transfer of DNA from *E. coli* to *P. polymyxa* by conjugation has been previously reported (Bezzate *et al.*, 2000), which offers the possibility of creation of targeted mutations in the chromosome of PKB1 strain using the Redirect© PCR targeting system originally developed for use with *Streptomyces coelicolor* (Gust *et al.*, 2003). In this strategy, a chromosomal gene cloned in a cosmid is replaced in *E. coli* by a PCR-amplified antibiotic resistance cassette that is selectable both in *E. coli* and in *Streptomyces*. The inclusion of an *oriT*$_{\mathrm{RK2}}$ site in the disruption cassette allows the mutant cosmid to be conjugated from *E. coli* into *Streptomyces* with the action of a non-transmissible plasmid pUZ8002. Theoretically, this protocol should be straightforward to extend its use to *P. polymyxa* because most genetic manipulations can be done in *E. coli* cells. However, some modifications were still required before the same technology could be applied to *P. polymyxa*.

3.5.4.1 Adaptation of the PCR-targeting protocol for *P. polymyxa*

In order to verify that intergeneric conjugation mediated by *oriT*$_{\mathrm{RK2}}$ and pUZ8002, provided as part of the Redirect© system, would occur between *E. coli* and *P. polymyxa*, the *E. coli-Paenibacillus* shuttle vector pC194-*apra* was constructed by ligation of the *Staphylococcus* plasmid pC194, which is able to replicate in *P. polymyxa* and imparts chloramphenicol resistance, into the *Hin*dIII site of the *E. coli* plasmid pIJ773, which has the Apra^R*oriT* disruption cassette and is supplied with the Redirect© PCR targeting materials. The plasmid pC194-*apra* now contains an origin of transfer (*oriT*$_{RK2}$), as well as autonomous origins of replication for both *E. coli* and *P. polymyxa*, and thus could be introduced into PKB1 strain by intergeneric conjugation. However, no Cm^R exconjugants were obtained on agar plates containing both chloramphenicol, to select for the plasmid pC194-*apra*, and nalidixic acid, to counterselect against *E. coli* donors in conjugation procedures of the original Redirect© protocol. In the meantime, the plasmid pC194-*apra* was transformed into PKB1 cells by electroporation as a control. *P. polymyxa* transformation mixtures were spread on agar plates containing chloramphenicol with or without nalidixic acid, but Cm^R transformants only occurred in the absence of nalidixic acid. The inhibitory effect of nalidixic acid on *P. polymyxa* was further confirmed by the observation that wild-type PKB1 cells were also sensitive to this chemical. Therefore, nalidixic acid could not be used with *P. polymyxa* and was replaced by polymyxin B sulfate, which exhibits antimicrobial activity mainly against Gram-negative bacteria, in conjugation experiments. The susceptibility of wild-type PKB1 to polymyxin B was tested and it was proved to be effective at killing *E. coli* but harmless to wild-type *P. polymyxa* PKB1 at 25 μg/ml.

Plating *P. polymyxa* conjugation mixtures on agar plates containing polymyxin B and chloramphenicol yielded many Cm^R colonies. To verify that these Cm^R colonies are true *P. polymyxa* exconjugants harboring the shuttle plasmid pC194-*apra*, and not spontaneous chloramphenicol resistant mutants or residual *E. coli* colonies, one Cm^R colony was picked and streaked on PDA plates to test for the ability to sporulate. The endospore-forming *P. polymyxa* colonies confirmed by microscopy were grown in liquid medium for plasmid preparation. The plasmid DNA isolated from *P. polymyxa* cells was first examined by agarose gel electrophoresis (Figure 3.5.5A), and then was used to transform *E. coli* cells. The plasmid purified from the apramycin and chloramphenicol resistant *E. coli* transformants was proved to be the plasmid pC194-*apra* by restriction analysis (Figure 3.5.5B), thus confirming that the shuttle plasmid pC194-*apra* could be transferred from *E. coli* to *P. polymyxa* PKB1 by intergeneric conjugation.

Initially it seemed that the Apra^R*oriT* cassette (from pIJ773) alone might be suitable for PCR-targeted mutation in *P. polymyxa*, since the *aac(3)IV* gene carried within the Apra^R*oriT* disruption cassette appeared to be a useful antibiotic marker that conferred apramycin resistance on *P. polymyxa* (Section 3.5.1). However, upon transfer of the shuttle plasmid pC194-*apra* into *P. polymyxa* PKB1 by conjugation, no Apra^R exconjugants could be recovered in the presence of both apramycin and polymyxin B. PKB1 transformants containing the plasmid pC194-*apra* were also found to be sensitive to this combination of antibiotics, although their growth was unaffected in the presence of either apramycin or polymyxin B alone. The lethal effect on *P. polymyxa* caused by combination of apramycin with polymyxin B is not fully understood, but it prevents use of the *aac(3)IV* gene as a selectable marker in conjugation experiments. Hence, a new plasmid construct, pJL9, carrying a hybrid gene disruption cassette containing *aac(3)IV*, *oriT*$_{RK2}$, and *cat* (from pC194) was prepared (Section 2.4.8), allowing for independent selection of Apra^R (encoded by *aac(3)IV*) in *E. coli* and Cm^R (encoded by *cat*) in *P.*

polymyxa. Like the disruption cassettes used in the *Streptomyces* system, the new Apra^{R}Cm^R*oriT* cassette employs the same 19-nt and 20-nt priming sequences for PCR amplification, and contains the same FRT sites for FLP recombinase-mediated excision (Figure 3.5.6).

3.5.4.2 Disruption of the putative peptide synthetase gene *orf7* on cosmid Col-19

In order to confirm the involvement of the *orf7* gene fragment identified on cosmid Col-19 in fusaricidin production, *orf7* mutants were created by using the PCR targeting mutagenesis protocol adapted for use in *P. polymyxa*, and the effect of this mutation on the antifungal activity against *L. maculans* was assessed. A segment (~2.4 kb) internal to *orf7* was targeted for PCR primer design. Amplification of the Apra^{R}Cm^R*oriT* cassette using primers JRL15-RD and JRL16-RD resulted in accumulation of the expected ~2.5 kb DNA fragment (Figure 3.5.7), which was subsequently transformed by electroporation into *E. coli* BW25113 carrying both pIJ790 expressing λ Red recombinase, and the *orf7*-bearing cosmid Col-19. With selection for Apra^R transformants, a mutagenized form of the Col-19 cosmid in which the internal 2.4-kb segment of *orf7* was deleted and replaced by the Apra^{R}Cm^R*oriT* cassette, was obtained through λ Red-mediated homologous recombination between the linear PCR product and the cosmid (Figure 3.5.8). The desired gene disruption in the mutant Col-19 cosmid was verified by restriction analysis and PCRs (see details below). The mutant Col-19 cosmid carrying the disrupted *orf7* allele was then passed through the nonmethylating *E. coli* host strain ET12567, and mobilized *in trans* by the non-transmissible plasmid pUZ8002 into PKB1 via intergeneric conjugation (Figure 3.5.9).

Among the resultant Cm^R exconjugants, four independent mutants (A1, A4, A5, and A6) were isolated which gave the expected size fragment (697 bp) in PCR tests using the *aac(3)IV*–specific primers SEJ18 and SEJ19 (Figure 3.5.10A). The second PCR reaction using SEJ19 primer with a flanking locus-specific primer JRL14 showed that all four mutants had new junctions and correct integration of the disruption cassette on their chromosomes (Figure 3.5.10B). A third reaction was performed with JRL14 primer and a locus-specific primer JRL30 to test for loss of the 2.4-kb segment targeted for replacement by the disruption cassette. As shown in Figure 3.5.10C, no PCR products were obtained from mutants A4 and A6, but the other two mutants A1 and A5 produced a locus-specific fragment (984 bp) that was expected to be amplified from PKB1 chromosome only, indicating the presence of a wild-type copy of *orf7* in A1 and A5 mutants.

Genomic DNA from the wild-type PKB1 strain and the four *orf7* mutants was also examined by Southern analysis to confirm the nature of the mutations. When the *cat*-containing *Bgl*II fragment from the plasmid pJL1 (Section 3.5.3) was used as a *cat*-specific probe, a 2.3-kb *Pst*I fragment hybridizing to the probe was detected in the four mutant samples, but not in the wild type, consistent with the expected gene replacement by the Apra^{R}Cm^R*oriT* cassette in all *orf7* mutants (Figure 3.5.11B-1). However, when the same blot was stripped and re-probed with an *Pst*I fragment internal to *orf7*, which encompasses the region replaced by the Apra^{R}Cm^R*oriT* cassette, a hybridizing band of the expected size (2.85 kb) was seen in the wild-type sample but also in the A1 and A5 mutants (Figure 3.5.11B-2). No hybridization to the probe was seen in lanes containing genomic DNA from the A4 and A6 mutants. Taken together, these findings suggest that disruption of *orf7* by the antibiotic resistance cassette in mutants A4 and A6 resulted from gene replacement via double crossover within *orf7*. However, complete sequence analysis of Col-19 revealed the presence of discontinuous genomic DNA fragments both upstream and downstream of *orf7* (Section 3.2.4). This finding leaves open the possibility that homologous recombinations may have occurred in the upstream and/or downstream non-*orf7* sequences, leading to deletion of other genes flanking *orf7* from the chromosome (see Section 3.6.1 for further discussion). In the other two *orf7* mutants A1 and A5, the entire mutant cosmid presumably integrated into the chromosome by single crossover, resulting in both a wild-type and a mutant copy of *orf7* present on their chromosomes.

3.5.4.3 Phenotype of the *orf7* mutants

The *orf7* disruptants were cultivated in sporulation medium PDB-soy, along with wild-type PKB1, to assess fusaricidin production. After 72 h, methanol extracts of the cell and spore pellets, the culture supernatant, and the entire culture were tested for antifungal activity against *L. maculans* (Figure 3.5.12). When compared to wild-type PKB1, mutants A4 and A6 had completely lost the ability to produce antifungal material. In contrast, the two single crossover mutants gave bioassay results indistinguishable from the wild type (data not shown for A1). Whether these mutants reverted to wild type upon growth in the absence of antibiotic selection, or whether production was due to the wild-type copy of the *orf7* gene remaining on their chromosomes was not determined.

Partially purified methanol extracts of wild-type and *orf7* mutant A4 cultures grown in PDB-soy medium were also analyzed by HPLC. This work was kindly performed by

Dr. Jensen and Annie Wong, a technician in our lab. Here, I present their HPLC results in support of my overall conclusions. Although no authentic standards are available for the various analogs of fusaricidins, extracts from the wild-type culture showed a series of A220 peaks eluting between 15 and 17.5 min upon reversed-phase chromatography, and fractions corresponding to these peaks showed bioactivity against *L. maculans*, consistent with the presence of fusaricidins. HPLC analysis of corresponding extracts from the A4 mutant showed a simpler profile with several peaks missing in this area, and no bioactivity was detected in these or any other fractions (Figure 3.5.13). On this basis, it was concluded that the A4 mutant does not produce detectable fusaricidins and therefore the putative peptide synthetase gene *orf7* is essential for fusaricidin biosynthesis.

When wild-type *P. polymyxa* PKB1 is grown on PDA or other carbohydrate-rich media, colonies exhibit a thick capsular layer of slime, presumably due to production of extracellular levan (Han, 1989). Interestingly, when the *orf7* mutants A4 and A6 were grown on PDA plates, the colonies produced noticeably larger amount of capsular slime than was observed for either wild-type PKB1 or the single crossover mutants A1 and A5. This observation may indicate a relationship between polysaccharide synthesis and antibiotic production in *P. polymyxa*.

3.6 Analysis of the fusaricidin biosynthetic gene cluster and the flanking regions

The two overlapping fosmids (4G9 and 6D11) isolated from the pSMART-FOS genomic library of *P. polymyxa* PKB1 (Section 3.4) yield a contiguous DNA sequence of approximately 48 kb (Appendix C), covering the entire fusaricidin biosynthetic gene cluster (*fus*) and flanking regions. The *fus* gene cluster spans 32.4 kb and includes 8 ORFs. The organization and assigned functions of these ORFs are shown in Figure 3.6.1 and Table 3.5, respectively. More than 6 kb of DNA from each flank of the *fus* cluster was also sequenced. The boundaries of the *fus* cluster are predicted based on the nature of the gene products encoded by the *fus* cluster, and on gene disruption analyses. Genes identified in the biosynthetic gene cluster include those necessary for assembly of the peptide backbone, and synthesis and incorporation of the lipid moiety of fusaricidin, but genes for regulation, export, and resistance were not detected in the region.

3.6.1 Modular organization of the fusaricidin synthetase

The *fus* gene cluster includes a large ORF of about 23.7 kb, encoding a protein of 7,908 amino acids with a molecular mass of 888,101 Da. The predicted gene product shows highest similarity to bacitracin synthetase (BacC) from *Bacillus licheniformis* (GenBank accession no. AAC06348). Sequence analysis revealed that the partial ORF *orf7* identified on cosmid Col-19 corresponds to the 5' end of this NRPS gene. Therefore, the complete ORF identified in the *fus* cluster was designated as the fusaricidin synthetase gene, *fusA*, and accordingly, the fusaricidin-nonproducer mutants A4 and A6 created by disruption of *orf7* (Section 3.5.4) are referred to as *fusA* mutants in the rest of this thesis. To ensure that targeted mutation of *fusA* occurred with no unintended deletion of neighboring genes, Southern analysis could have been performed, in which genomic DNA from wild-type PKB1 and the A4 and A6 mutants was digested with restriction enzymes that cut outside of *fusA* and generate large-size restriction fragments, and then the fractionated and blotted DNA was probed with DNA fragments corresponding to regions flanking *fusA*. Alternatively, a wild-type copy of *fusA* carried on a plasmid could have been introduced into the *fusA* mutant (A4 or A6) to see whether fusaricidin production could be restored. However, such a complementation test would be made difficult by the very large size of *fusA* and lack of its own promoter for gene expression in *P. polymyxa*.

The translation of *fusA* appears to start with an ATG codon at nt 17,682, located 8 bp downstream of a putative ribosome-binding site (AGGAG) (Appendix C). By comparison with other peptide synthetases, six functional modules were identified within the deduced amino acid sequence (FusA) of *fusA* (Figure 3.6.2). Each module contains the three common catalytic domains of NRPSs, an adenylation domain, a thiolation domain, and a condensation domain. The last module terminates with a C-terminal thioesterase domain that is presumably required for cyclization and release of the peptide product. Most of the highly conserved core motifs characteristic for individual domains of NRPSs are found in the modules of fusaricidin synthetase.

3.6.1.1 Condensation (C) domains

The fusaricidin synthetase FusA contains five typical C domains that presumably catalyze peptide bond formation between two adjacent amino acid residues of fusaricidin. An additional C domain was detected at the N-terminus of the first module of FusA, which shows the highest similarity (approximately 40% identity) to the first C domain of the surfactin synthetase SrfA-A (GenBank accession no. BAA02522) (Figure 1.5.1), responsible for the incorporation of the first three amino acids into the lipopeptide surfactin, but only around 23% identity to the remaining five C domains in FusA. This can be explained by the observation that C domains are not normally present in the initial modules of NRPSs, except for lipopeptides, a subgroup of nonribosomally synthesized peptides whose peptide chain is N-terminally attached to a fatty acid (Section 1.5). The presence of an N-terminal C domain (C_N) preceding an A and a T domains in the initiation module, as seen in the first module of FusA, has been reported for many characterized lipopeptide synthetase systems from bacilli, actinomycetes, and pseudomonads, such as the surfactin (Cosmina *et al.*, 1993), daptomycin (Miao *et al.*, 2005), and arthrofactin (Roongsawang *et al.*, 2003) synthetases. These N-terminal C domains are presumed to be required for attachment of the lipid moiety as an early step of lipopeptide biosynthesis. Since they catalyze the coupling of fatty acids, rather than amino acids, to the first amino acid of the peptide chain, these initial C domains may share features not found in regular C domains, and so greater similarity is seen within this group than with other internal C domains (Figure 3.6.3). The primary structure of fusaricidin shows an N-terminal Thr residue acylated with a β-hydroxy fatty acid, as is also the case for surfactin. Thus, the high homology between the first C domain in FusA and that in SrfA-A further supports the proposal that the peptide synthetase encoded by *fusA* is responsible for fusaricidin biosynthesis.

Moreover, the remaining five C domains in FusA can be subdivided into two groups according to their locations. They are C_D domains (in FusA-C3, FusA-C5, and FusA-C6) that are preceded by an epimerization domain and accept an upstream D-amino acid at the donor site, and C_L domains (in FusA-C2 and FusA-C4) that are located immediately downstream of a T domain and accept an upstream L-amino acid at the donor site. Distinct differences between the C_D and C_L domains and similarity within the same domain type were seen in the amino acid sequences of the conserved motifs (Figure 3.6.3). This may reflect the stereoselectivity of C domains towards the C-terminal amino acid of the growing peptidyl chain in the course of the condensation reaction (Section 1.4.1.3).

3.6.1.2 Epimerization (E) domains

In most peptide synthetases produced by Gram-positive bacteria, modules that incorporate D-configured residues contain an additional domain responsible for epimerization, found downstream of the T domain. An L-amino acid is activated and then the E domain catalyzes L to D racemization of the thioester-bound amino acid (Section 1.4.1.5). Modifying epimerization (E) domains were detected at the C-terminal ends of the second, fourth and fifth modules of FusA, consistent with the incorporation of D-amino acids in these positions of fusaricidin peptides. However, the sixth module FusA-M6, corresponding to D-Ala, lacks such an E domain, suggesting that a different mechanism is adopted for D-Ala incorporation in fusaricidin biosynthesis.

Two distinct mechanisms of D-amino acid incorporation by a module without an E domain have been encountered so far. In the lipo-undecapeptide arthrofactin, for example, there are no E domains detected in any of the three arthrofactin synthetases, although seven of the eleven amino acids are in the D-configuration. Biochemical analyses demonstrated that A domains in modules corresponding to D-amino acids were specific for activation of L-isomers, and epimerase activity was provided by a new type of C domain with dual epimerization and condensation functions, located immediately downstream of the T domain acylated with the amino acid undergoing epimerization

(Balibar *et al.*, 2005; Roongsawang *et al.*, 2003). It was also suggested that the occurrence of D-amino acids in lipopeptides syringomycin (Guenzi *et al.*, 1998), syringopeptin (Scholz-Schroeder *et al.*, 2003), enduracidin (Yin and Zabriskie, 2006), and ramoplanin (Walker *et al.*, 2005) probably involves the same mechanism as demonstrated for arthrofactin.

The second, but very rare strategy for incorporation of D-amino acids without the action of an E domain, involves the direct activation of D-isomers by the A domains. Initially, only two instances of this mechanism are known, in cyclosporin (Dittmann *et al.*, 1994; Weber *et al.*, 1994) and HC-toxin synthetases (Scott-Craig *et al.*, 1992; Walton, 1987), both from fungal systems. Although these NRPSs have been reported to incorporate D-Ala directly, no biochemical studies using purified recombinant A domains have been conducted. However, both gene clusters do encode independent alanine racemases to generate the required D-Ala (Cheng and Walton, 2000; Hoffmann *et al.*, 1994). As this thesis was being written, another example of a D-Ala-specific A domain of bacterial origin was reported for the leinamycin synthetase (Tang *et al.*, 2007), a hybrid nonribosomal peptide-polyketide synthetase produced by *Streptomyces atroolivaceus* (Section 1.6). Cloning and sequencing of the leinamycin biosynthetic gene cluster revealed a novel type of NRPS initiation module, composed of two isolated proteins: LnmQ (adenylation protein) and LnmP (thiolation protein) (Figure 1.6.2). It was demonstrated biochemically that LnmQ directly activates D-Ala and loads it onto LnmP to initiate the biosynthesis of leinamycin (Tang *et al.*, 2007).

Since the amino acid incorporated by FusA-M6 occupies the C-terminal position of the fusaricidin peptide, this excludes the possibility that epimerization of an L-Ala residue is catalyzed by a downstream dual C/E domain, as seen in the case of arthrofactin synthetases. Therefore, either free D-Ala may be directly selected and activated by the A domain of FusA-M6, as was shown for leinamycin biosynthesis, or possibly an external racemase may be involved in transformation of thioester-bound L-Ala to D-Ala.

3.6.1.3 Thiolation (T) domains

The six thiolation domains in FusA can be divided into two groups according to their conserved motifs. Modules FusA-M2, FusA-M4 and FusA-M5 are associated with E domains, and in these modules, the T domains upstream of E domains all contain a core sequence of LGG**D**SIK. It has been demonstrated that the aspartate residue (in bold) in front of the conserved serine residue in T domain core motifs is essential for proper interaction between the T domain and downstream E domain, and subsequent racemization of the thioester bound L-amino acid to its D-isomer (Linne *et al.*, 2001). In contrast, the core sequence of the T domain found in FusA-M6 (LGG**H**SL) matches that of T domains not associated with E domains (xGG**H**SL), such as those found in the first and third modules. This finding is consistent with the absence of an E domain in the termination module of FusA. Furthermore, significant differences between these two groups of T domains were observed in the downstream regions of the core motifs (Figure 3.6.4), which may reflect the requirement of proper positioning of T domain-bound amino acid at the E domain during substrate epimerization.

3.6.1.4 Adenylation (A) domains

By comparison with the amino acids lining the substrate-binding pocket in the Phe-activating domain of the gramicidin S synthetase GrsA, a "nonribosomal code" specifying important residues involved in substrate recognition by A domains has been defined by two independent groups (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999) (Section 1.4.1.1). The corresponding amino acid residues were examined in the six A domains of FusA, and the substrate specificity of each A domain was predicted (Table 3.6).

The A domain in the first module FusA-A1 incorporating L-Thr, and the one in the fourth module FusA-A4 incorporating D-*allo*-Thr, both share the same signature sequence as the Thr-activating domain from fengycin synthetase FenD (GenBank accession no. CAA09819).

The substrate recognition sequence of FusA-A2 is most similar to that of the Val-activating A domain from surfactin synthetase SrfA-B (GenBank accession no. BAA08983), which also activates isoleucine to a lesser extent (Elsner *et al.*, 1997). The presence of tryptophan at position 278 of the specificity code was suggested to be important for recognition of the smaller side chain of valine as compared to that of isoleucine (Challis *et al.*, 2000), and a tryptophan residue is found at this position in FusA-A2. The incorporation of D-Val, D-Ile, or D-*allo*-Ile as substrates in the second position of fusaricidins indicates an even broader substrate specificity, possibly due to substitutions of Ile299→Leu and Gly322→Cys in FusA-A2 compared to SrfA-B.

FusA-A3 contains a signature sequence most closely resembling the 3-hydroxy-L-Tyr-activating domain from chloroeremomycin synthetase CepB (GenBank accession no. CAA11795). However, considerable similarity to the Phe-activating domains from TycA and GrsA synthetases (GenBank accession no. AAC45928, and CAA33603, respectively) was also noted in the A4-A5 regions that define most of the substrate-binding pockets in A domains. In addition, the specificity-conferring residues identified in FusA-A3 are mainly hydrophobic. Together, these findings may indicate an overall relaxed specificity towards aromatic and hydrophobic amino acid residues at the third position, as seen in the fusaricidin variants.

The specificity code of FusA-A5 matches that of the Asn-activating A domain from tyrocidine synthetase TycC (GenBank accession no. AAC45930). Replacement of D-Asn by another polar residue D-Gln at the fifth position in several fusaricidin variants indicates that this conservative substitution is tolerated by the corresponding A domain. Although the same substrate recognition sequence (DLTKIGEVG) is found in the Asn-activating A domain derived from the *fus-asn* gene fragment, previously subcloned from the PKB1 chromosome (Section 3.2.1), FusA-A5 is not identical to *fus-asn* in the nucleotide sequences. This further confirmed that *fus-asn* should be part of a separate NRPS gene cluster responsible for biosynthesis of Asn-containing peptide metabolites, other than fusaricidins, in *P. polymyxa* PKB1.

The amino acid substrate for FusA-A6 cannot be predicted because its signature sequence shows no similarity to A domains with assigned specificities, including those activating L- and D-Ala (Figure 3.6.5). Ala-specific A domains, whether for L- or D-isomers, show relatively weak similarity, perhaps because the small size of the alanine sidechain allows greater variability in the residues lining the Ala-binding pocket to be tolerated (Challis *et al.*, 2000). With the exception of FusA-M6, the predicted substrates for the A domains of each module of fusaricidin synthetase correspond to the residues found at the respective positions of the fusaricidin peptides.

3.6.2 Determination of the substrate specificity of the A domain in the sixth module of FusA

In order to determine the substrate specificity of FusA-A6, a DNA fragment corresponding to the adenylation domain of FusA-M6 was amplified from fosmid 6D11, and cloned into pET-19b. Based on the previously described A domain borders (Mootz and Marahiel, 1997), the N-terminal end of the FusA-A6 domain was set at 101 aa upstream from the core motif A2 (LKAGGA), and the C-terminal end at 17 aa upstream from the core motif T (LGGHSL) (Figure 3.6.6). After expression in *E. coli*, a soluble protein with the predicted molecular mass of 64 kDa was obtained and purified by Ni^{2+} affinity chromatography (Figure 3.6.7). Enzymatic activity of the purified recombinant FusA-A6 protein was determined by an ATP-PPi exchange assay with various amino acids as substrate. From the results obtained (Figure 3.6.8), it was evident that FusA-A6 is specific for activation of D-Ala, with very low tolerance for L-Ala (less than 1% of D-Ala), or any other amino acids (1-4% of D-Ala).

In previous studies of cyclosporin and HC-toxin synthetases (Dittmann *et al.*, 1994; Scott-Craig *et al.*, 1992; Walton, 1987; Weber *et al.*, 1994), the specificity for D-Ala activation by respective A domains was predicted based on the colinearity rule of NRPSs. However, the D-Ala-dependent ATP-PPi exchange reactions that were performed in both fungal systems used entire peptide synthetase proteins and therefore could not exclude the possibility that external racemases were co-purified from the producer organisms

(Cheng and Walton, 2000; Walton, 1987; Zocher *et al.*, 1986). The data from the present study, as well as the leinamycin synthetase system (Tang *et al.*, 2007), clearly show that a recombinant A domain, from a peptide synthetase of prokaryotic origin, exclusively recognizes and activates a D-amino acid as the substrate. While the cyclosporin and HC-toxin biosynthetic gene clusters each encode a distinct alanine racemase catalyzing conversion of L- to D-Ala (Cheng and Walton, 2000; Hoffmann *et al.*, 1994), an alanine racemase specific for the *fus* cluster or the leinamycin biosynthetic gene cluster was not encountered. However, alanine racemases are widely present in bacteria where they provide D-Ala essential for cell wall synthesis.

3.6.3 N-terminal lipidation of fusaricidin

Fusaricidins belong to a group of nonribosomal peptides that are modified at the N-terminus with a fatty acid. The lipid side chains of lipopeptide antibiotics are believed to play an important role in their interactions with their cellular targets, such as cell membranes. The predicted gene products of six ORFs upstream of *fusA* show homology to enzymes involved in either fatty acid synthesis (*fusB*, *fusC*, *fusE*, *fusF* and *fusG*) or amino acid biosynthesis (*fusD*) (Table 3.5), and accordingly, are likely to synthesize and/or modify, and activate a fatty acid precursor for attachment to the amino group of the N-terminal L-Thr residue of fusaricidin. The acyl-CoA ligase encoded by *fusF* may couple coenzyme A (CoA) to an acyl group, which is then transferred to an acyl carrier protein (ACP) for elongation, possibly involving the gene products of *fusB, fusC* and *fusG*, although no gene encoding an ACP-like protein was found in the *fus* cluster. During this process, a β-OH group must be preserved or introduced into the growing fatty acid chain, and ultimately, the activated fatty acid must be further modified by addition of a guanidino group. Alternatively, perhaps an arginine residue is deaminated to 5-guanidino pentanoic acid, then activated and transferred to ACP for elongation by a process analogous to type II fatty acid biosynthesis. Acylation of the N-terminal amino acid with the modified fatty acid is presumably catalyzed by the initial condensation domain in the first module of FusA (Section 3.6.1.1). Further studies are needed to clarify the catalytic roles of these fatty acid genes upstream of *fusA* in fusaricidin biosynthesis.

3.6.4 The 3' boundary of the *fusA* gene cluster

At 1,141 bp downstream from the translational stop codon of *fusA* and oriented in the opposite direction, a 624-bp ORF encoding a putative protein of 207 amino acids was detected (Figure 3.6.1). This gene product shows the greatest similarity to proteins of the α/β hydrolase fold family (~45% identity) and has one conserved GxSxG motif typical of thioesterases. Although it shows no significant similarity to the C-terminal TE domain of FusA, it may catalyze hydrolysis of acyl groups inappropriately attached to the 4'-PP cofactors of fusaricidin synthetase, as have been demonstrated for other type II thioesterases (Section 1.4.1.4). Therefore, this ORF was designated as *fusTE*. DNA sequence further upstream of *fusTE* showed 99% identity to a *P. polymyxa* 16S rRNA gene and beyond that a 23S rRNA gene is apparent (Table 3.5). Therefore, *fusTE* was assigned as the 3' boundary of the *fusA* gene cluster even though no gene disruption studies were conducted to confirm its involvement.

3.6.5 The 5' boundary of the *fusA* gene cluster

3.6.5.1 Preparation of the *ymcC* mutant

About 10 kb of DNA sequence upstream of *fusG* was examined in order to identify additional genes potentially involved in fusaricidin production (Figure 3.6.1). An ORF (*ymcC*), starting 1,160 bp upstream of *fusG*, encodes a protein product of 177 aa with similarity to YmcC (from *Bacillus licheniformis* ATCC 14580), a protein with unassigned function. The counterpart of *ymcC* in the *Bacillus subtilis* 168 genome precedes a cluster of putative polyketide synthase genes. In order to determine whether *ymcC* plays a role in fusaricidin biosynthesis, *ymcC* mutants were created using a PCR-targeting mutagenesis method as previously described for preparation of *fusA* mutants, with a minor change. The primers JRL42-RD and JRL43-RD were used to amplify by PCR the Apra^{R}Cm^R*oriT* disruption cassette from the template plasmid pJL9. The λ Red expression plasmid pIJ790 could not be used with the *P. polymyxa* fosmid clones since both pIJ790 and the pSMART-FOS vector confer chloramphenicol resistance. Therefore, the *ymcC*-bearing fosmid 4G9 was transformed into *E. coli* strain BW25113 carrying pKD46. Like pIJ790, pKD46 expresses λ Red functions when induced by L-arabinose, but it contains the ampicillin resistance gene *bla* instead of the Cm^R marker, allowing selection in the presence of pSMART-FOS-derived fosmids. The PCR-generated disruption cassette was then introduced by electroporation into *E. coli* BW25113 harboring both pKD46 and 4G9. With selection for Apra^R transformants, the mutant fosmid 4G9:Δ*ymcC* was obtained where the entire *ymcC* gene was deleted and replaced by the disruption cassette. The mutagenized fosmid was introduced into the wild-type *P. polymyxa* PKB1 strain by intergeneric conjugation (Figure 3.6.9), and four *ymcC* mutants were identified among the Cm^R exconjugants by PCRs.

Two PCR reactions were carried out using the *aac(3)IV*–specific primers SEJ18 and SEJ19, and primer SEJ18 with a flanking locus-specific primer JRL45, respectively. All four mutants gave locus-specific fragments of the expected sizes, indicating that they have the Apra^{R}Cm^R*oriT* disruption cassette integrated at the correct locus on their chromosomes (Figure 3.6.10A). A third reaction, using two flanking locus-specific primers JRL44 and JRL45 to test for loss of the target gene, revealed that all *ymcC* mutants had the expected-size fragments specific for gene replacement by the disruption cassette, but mutants #3 and #12 also produced the *ymcC*-specific fragments, indicating the presence of a wild-type copy of *ymcC* in both mutants (Figure 3.6.10B). Together, these observations suggest that mutants #8 and #11 probably resulted from the replacement of *ymcC* with the gene disruption cassette via double crossover, while the other two mutants #3 and #12 contain both wild-type and mutant copies of *ymcC*, presumably resulting from integration of the entire fosmid into the chromosomes by a single homologous recombination.

The *ymcC* mutant (#8) that had been isolated and confirmed by PCR reactions was fermented in PDB-soy medium, along with the wild□type PKB1 strain. After 72 hours of growth, fusaricidin production in both cultures was assessed by bioassay with *L. maculans* as the indicator organism. Since the zones of inhibition produced by *ymcC* mutant and wild-type PKB1 are quite similar in size (Figure 3.6.11), disruption of the *ymcC* gene appeared to have no apparent effect on production of fusaricidin.

3.6.5.2 Preparation of the *ktp* mutant

ABC-type transporter systems were identified within several peptide synthetase gene clusters (Gaisser and Hughes, 1997; Mootz and Marahiel, 1997; Quigley *et al.*, 1993; Roongsawang *et al.*, 2003). Gene disruption of either the transporter or the synthetase in *Proteus mirabilis* resulted in the same swarming-defective phenotype (swarming appeared to be associated with peptide/polyketide production in this case), suggesting that the transporter proteins may play a role in the secretion of corresponding peptide products (Gaisser and Hughes, 1997). However, no genes encoding transporter-like proteins were found in the sequenced region downstream of *fusA*.

An ORF, oriented in the same transcriptional direction as *ymcC*, is located further 1,630 bp upstream (Figure 3.6.1). The predicted gene product of 445 aa shares sequence similarity with potassium uptake transporter proteins and so this ORF was designated *ktp*. To determine if this protein product was involved in export of fusaricidins, *ktp* mutants were generated by following the same PCR-targeting protocol as described in the preparation of *ymcC* mutants, except that the PCR primers JRL31-RD and JRL32-RD were used to amplify the Apra^{R}Cm^R*oriT* disruption cassette. By using three PCR reactions with *aac(3)IV*–specific primers and locus-specific primers, one *ktp* mutant T1 was identified, in which the wild-type copy of *ktp* was replaced by the Apra^{R}Cm^R*oriT* disruption cassette via double crossover recombination (Figure 3.6.12). However, this mutant strain still produced wild-type level of fusaricidins, as judged by bioassay of culture extracts against the indicator fungus *L. maculans* (Figure 3.6.11), thus indicating that the gene product of *ktp* is not required for export of, or self-resistance to, fusaricidins.

In the sequenced region further upstream from *ktp*, a variety of housekeeping genes were encountered (Table 3.5), and most of them have functional counterparts in *Bacillus* genomes. However, none of these ORFs encoded functions obviously related tfusaricidin biosynthesis. Therefore, the 5' boundary of the *fus* cluster is likely defined by *fusG*.[1][2]

115

---

[1] *A version of this chapter has been published. Li et al. 2007. Applied and Environmental Microbiology. 73: 3480-3489.*

[2] *A version of this chapter has been submitted for pulication. Li and Jensen 2007.*

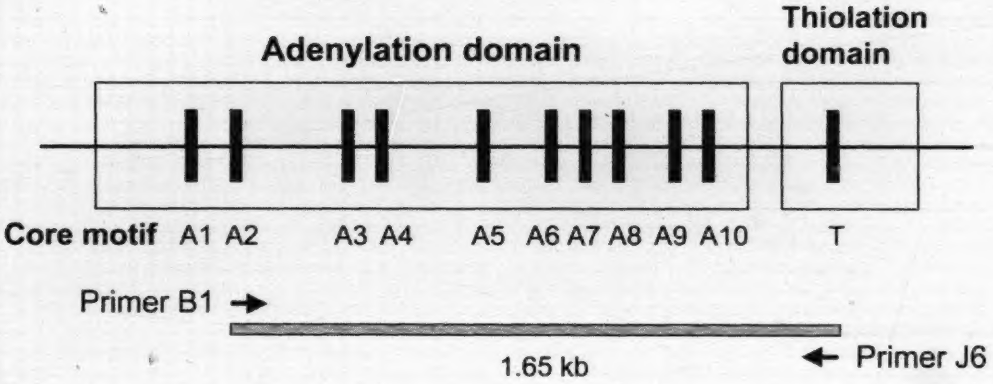

Adenylation domain
Thiolation domain
Core motif
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
T
Primer B1
1.65 kb
Primer J6

Figure 3.1.1 The PCR primers used to amplify fragments of the peptide synthetase genes from *P. polymyxa* PKB1. The oligonucleotide primers B1 and J6 were designed to recognize the conserved core motifs A2 and T, common to all peptide synthetases, respectively. Annealing of these primers to peptide synthetase genes would result in amplification of a DNA fragment of approximately 1,650 bp.

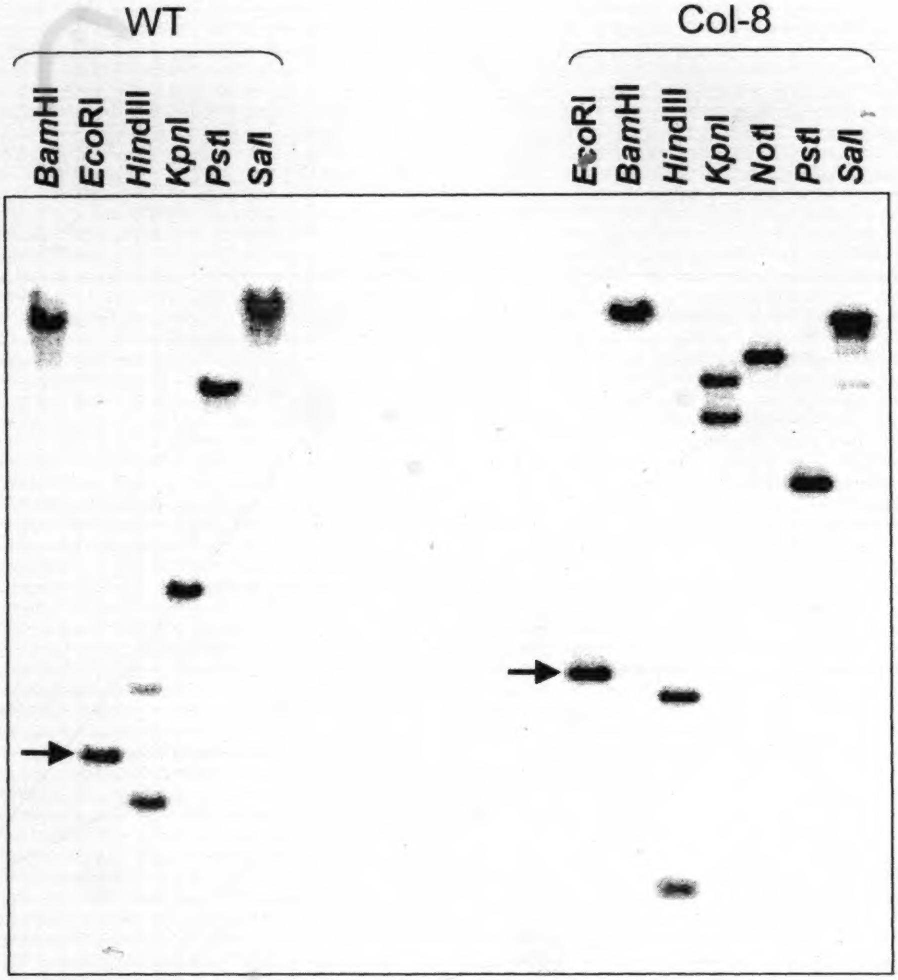

WT
BamHI
EcoRI
HindIII
KpnI
PstI
SalI
Col-8
EcoRI
BamHI
HindIII
KpnI
NotI
PstI
SalI

Figure 3.2.1 Southern analysis of cosmid Col-8 identified in the first screening of the SuperCos-1 genomic library of *P. polymyxa* PKB1. The cosmid DNA from Col-8 and genomic DNA from wild-type PKB1 (WT) were digested with various restriction enzymes before being subjected to Southern analysis. The membrane with blotted DNA was probed with the PCR-amplified fragment B1J6-10. The two *Eco*RI fragments from either Col-8 or genomic DNA, which hybridized to the probe, are indicated.

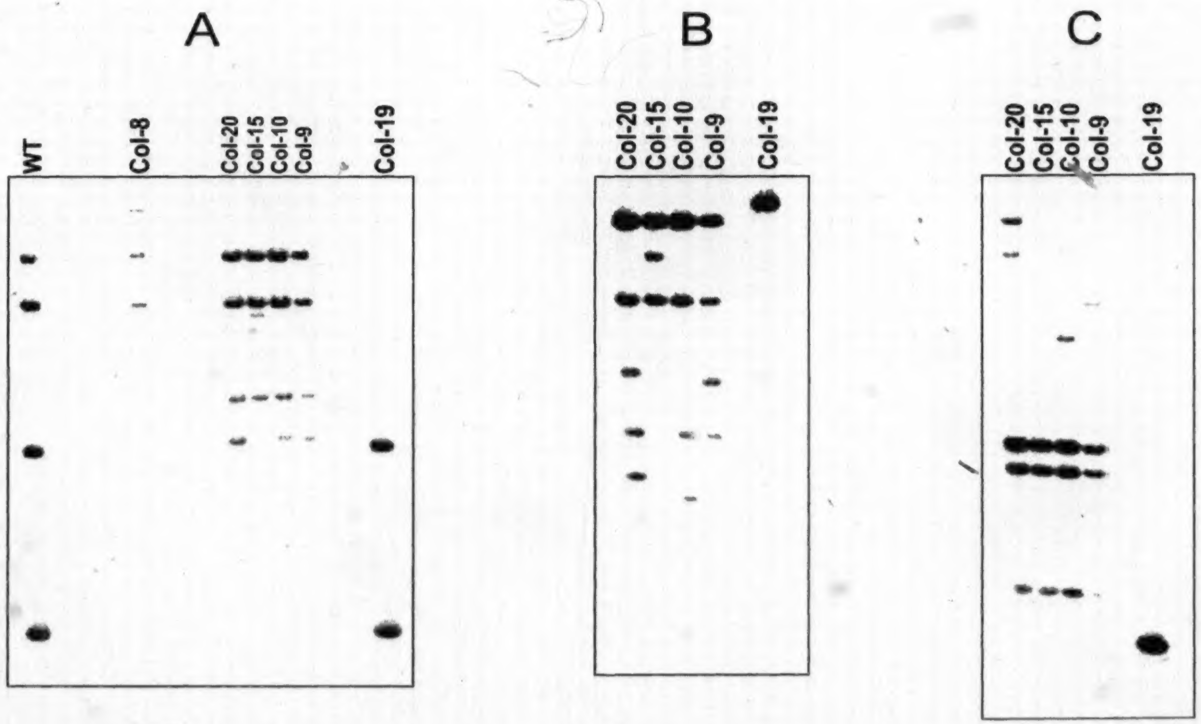

A
WT
Col-8
Col-20
Col-15
Col-10
Col-9
Col-19
B
Col-20
Col-15
Col-10
Col-9
Col-19
C
Col-20
Col-15
Col-10
Col-9
Col-19

Figure 3.2.2 Southern analysis of the positive cosmids identified in the second screening of the first SuperCos-1 genomic library of *P. polymyxa* PKB1. The cosmid DNA from Col-8, Col-9, Col-10, Col-15, Col-19 and Col-20, as well as genomic DNA from the wild-type PKB1 strain (WT), was digested with *Pst*I (A), *Eco*RI (B) and *Hin*dIII (C), respectively. The PCR-amplified fragment B1J6-17 was used as a probe in the Southern analysis.

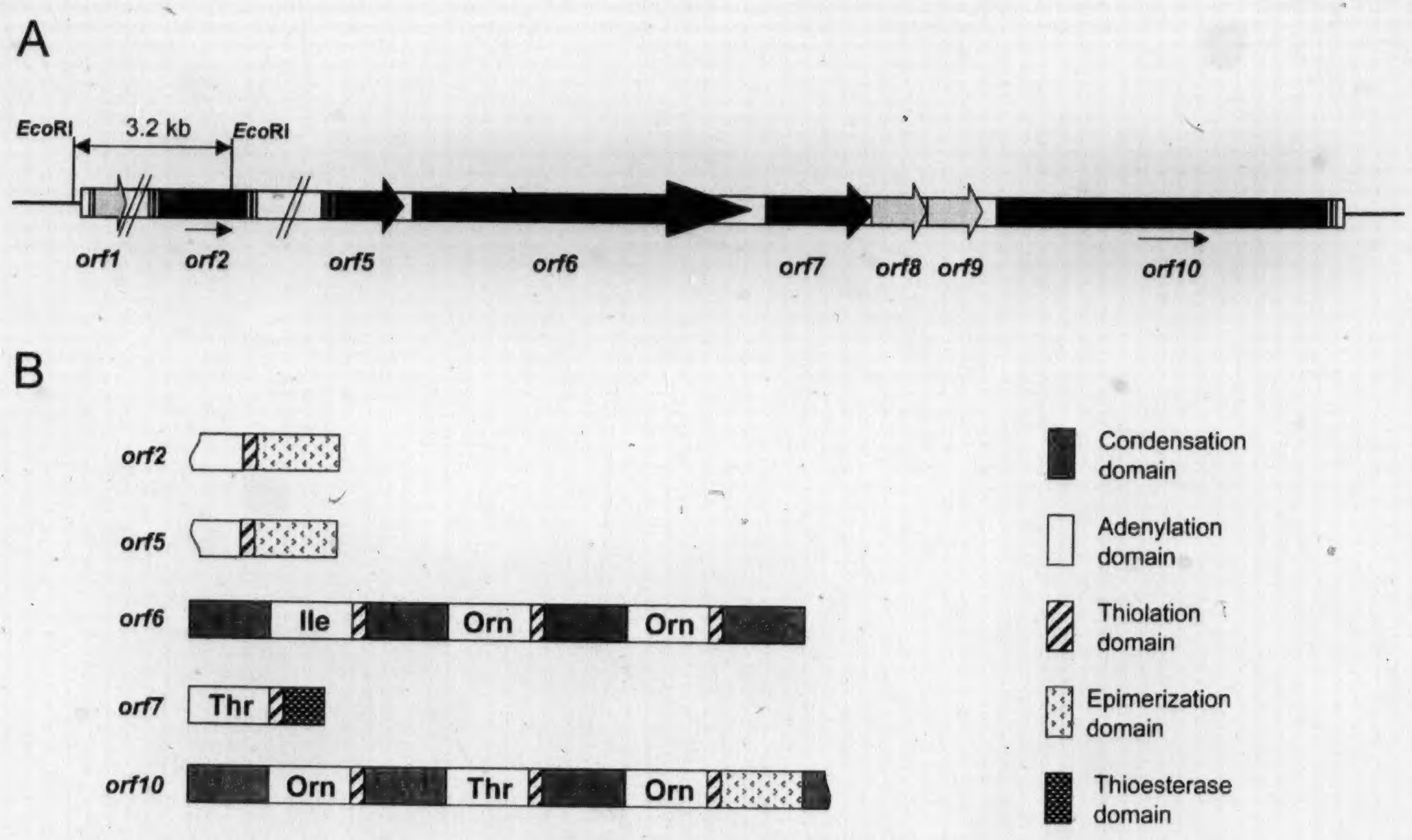
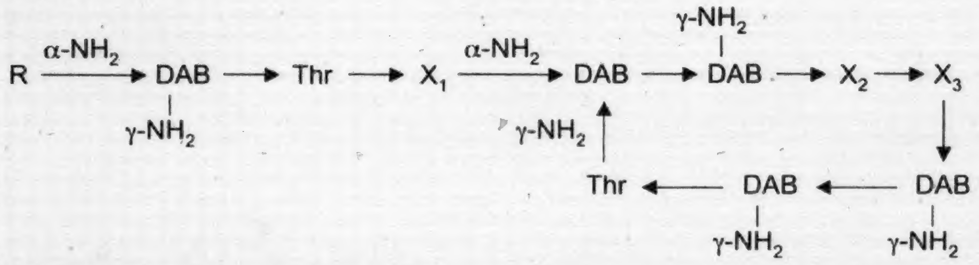

A
EcoRI
3.2 kb
EcoRI
orf1
orf2
orf5
orf6
orf7
orf8
orf9
orf10
B
orf2
orf5
orf6
Ile
Orn
Orn
orf7
Thr
orf10
Orn
Thr
Orn
Condensation domain
Adenylation domain
Thiolation domain
Epimerization domain
Thioesterase domain

Figure 3.2.3 ORF organization of the insert DNA from cosmid Col-8.

Figure 3.2.3 (continued) (A) The long white bar represents the insert of Col-8, and the fine line represents the SuperCos-1 cosmid vector. ORFs encoding peptide synthetases are represented by black arrows (or bars) and the other ORFs are shown in gray. The truncated ends of incomplete ORFs are shown by short vertical double lines, and their transcriptional directions are indicated by arrows. The long double lines represent possible discontinuities in the nucleotide sequence of cosmid insert. The 3.2-kb *Eco*RI fragment previously subcloned from Col-8 is also indicated. The figure is not drawn to scale. (B) Domain organizations of the peptide synthetases derived from *orf2*, *orf5*, *orf6*, *orf7*, and *orf10*. The incomplete adenylation domains identified within the deduced amino acid sequences of *orf2* and *orf5* do not contain enough sequence information to allow predication of their substrate specificities.



Polymyxin	R	X$_1$	X$_2$	X$_3$
Polymyxin B1	6-methyloctanoic acid	DAB	D-Phe	Leu
Polymyxin B2	6-methylheptanoic acid	DAB	D-Phe	Leu
Polymyxin E1 (Colistin A)	6-methyloctanoic acid	DAB	D-Leu	Leu
Polymyxin E2 (Colistin B)	6-methylheptanoic acid	DAB	D-Leu	Leu
Circulin A	6-methyloctanoic acid	DAB	D-Leu	Ile
Circulin B	6-methylheptanoic acid	DAB	D-Leu	Ile
Polymyxin S	6-methyloctanoic acid	D-Ser	D-Phe	Thr

[a] DAB: 2,4-diaminobutyric acid

[b] The fatty acid chain is attached to the α-amino group of the N-terminal DAB.

[c] Cyclization of the peptide occurs via an amide bond between the γ-amino group of DAB at the fourth position and the α-carboxyl group of the C-terminal threonine residue.

Figure 3.2.4 The structures of commonly used polymyxin-type antibiotics produced by *P. polymyxa* (Katz and Demain, 1977). All amino acids are in the L-configuration unless indicated otherwise.

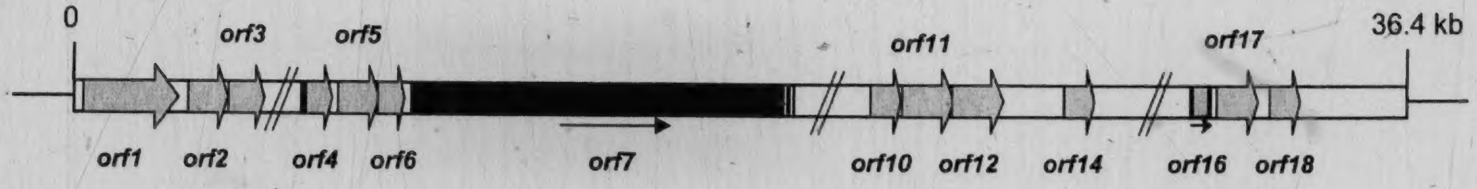

0
orf3
orf5
orf11
orf17
36.4 kb
orf1
orf2
orf4
orf6
orf7
orf10
orf12
orf14
orf16
orf18

Figure 3.2.5 ORF organization of the insert DNA from cosmid Col-19. The black arrow represents the peptide synthetase gene *orf7*, while the other ORFs are shown in gray. The truncated ends of incomplete ORFs are shown by short vertical double lines, and their transcriptional directions are indicated by arrows. The long double lines represent possible discontinuities in the nucleotide sequence of cosmid insert. The figure is not drawn to scale. See Table 3.4 for predicted functions of assigned ORFs.

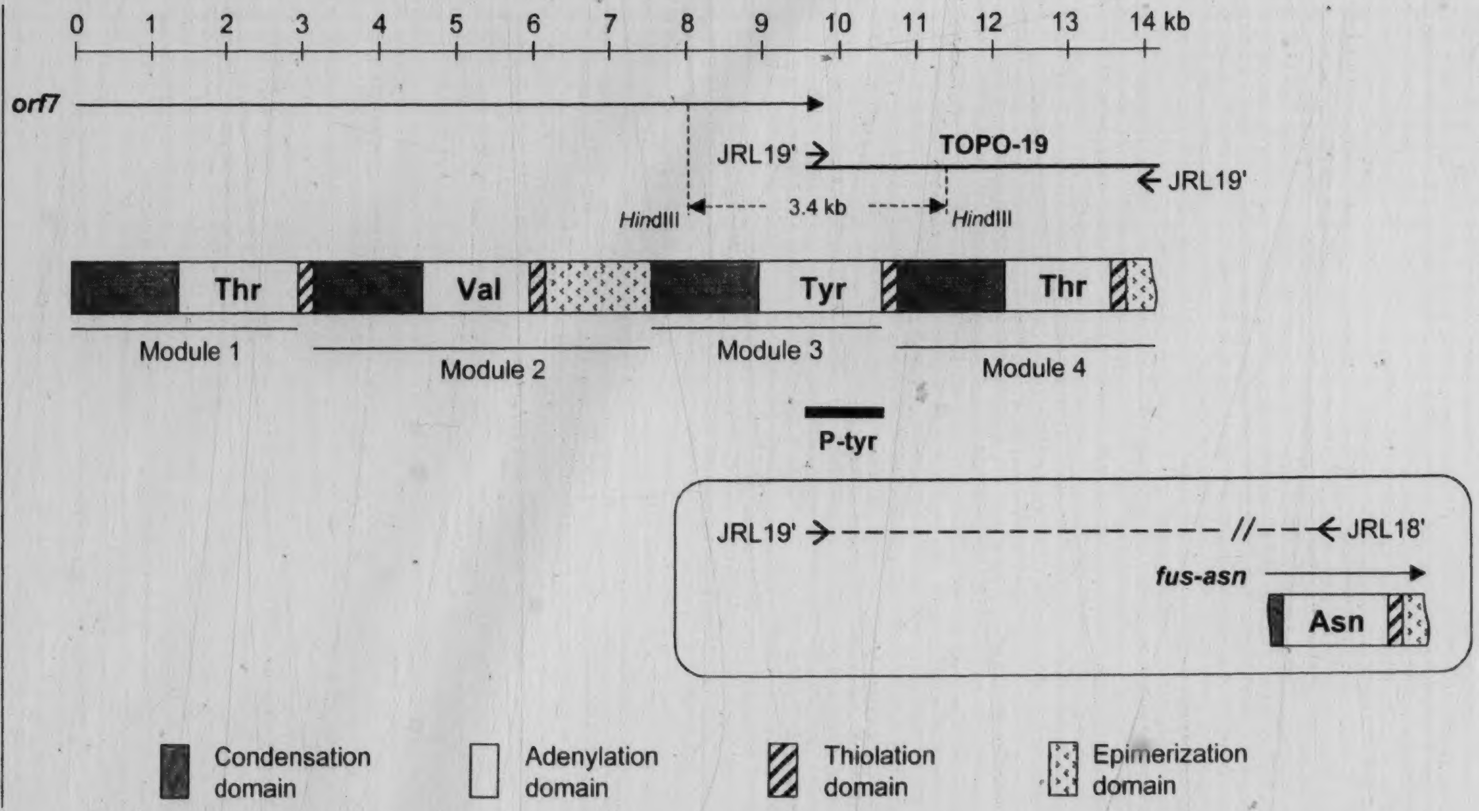

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14 kb
orf7
JRL19'
TOPO-19
JRL19'
HindIII
3.4 kb
HindIII
Thr
Val
Tyr
Thr
Module 1
Module 2
Module 3
Module 4
P-tyr
JRL19'
//
JRL18'
fus-asn
Asn
Condensation domain
Adenylation domain
Thiolation domain
Epimerization domain

Figure 3.2.6 Schematic representation of the peptide synthetase gene comprising *orf7* from Col-19 and the PCR-amplified fragment TOPO-19.

Figure 3.2.6 (continued) By using *P. polymyxa* PKB1 genomic DNA as template, the fragment TOPO-19 was amplified with primer JRL19', derived from the 3' end of *orf7*. Domain organizations of the peptide synthetase deduced from the assembled DNA sequence, and from the *fus-asn* fragment are illustrated. The predicted substrates of the adenylation domains are indicated. The dashed line represents the assumed PCR product amplified by using primer JRL19', and primer JRL18' derived from the A-domain in the deduced amino acid sequence of *fus-asn*. The P-tyr probe used for screening the fosmid genomic library of *P. polymyxa* PKB1 and the 3.4-kb *Hin*dIII fragment hybridizing to the P-tyr probe (Figure 3.4.1) are also shown.

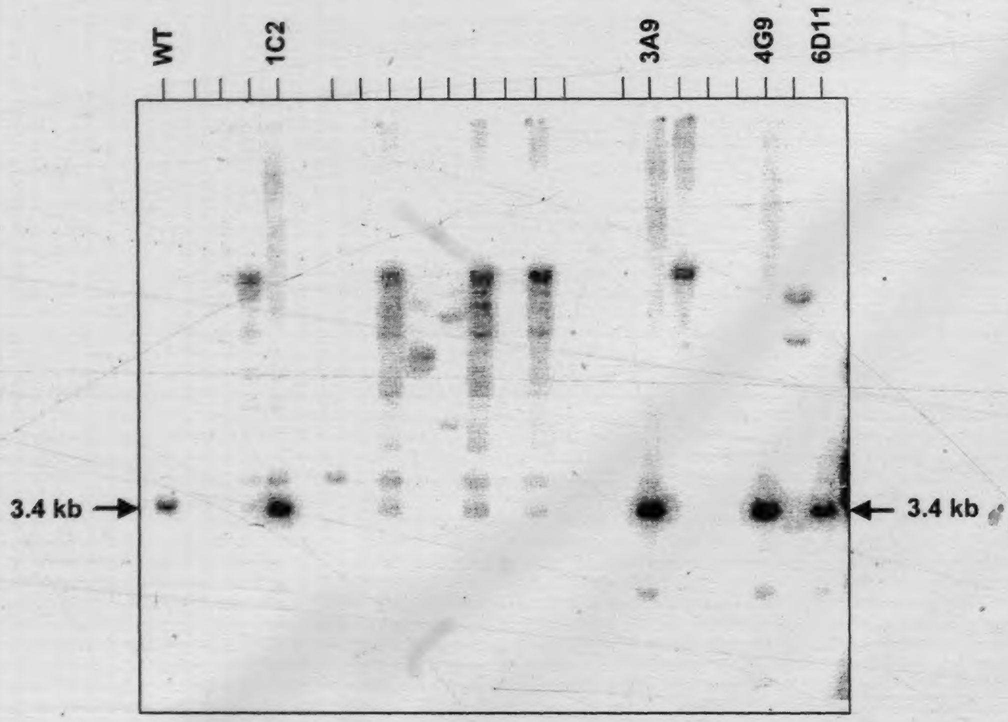


Figure 3.4.1 Southern analysis of the hybridizing fosmids identified in the pSMART-FOS genomic library of *P. polymyxa* PKB1. The fosmid DNA from twenty-one positive clones and genomic DNA of wild-type PKB1 (WT) were digested with *Hin*dIII before being subjected to Southern analysis. The membrane was hybridized with the probe P-tyr, corresponding to the Tyr-activating adenylation domain (Figure 3.2.6). The four fosmids, bearing the 3.4-kb *Hin*dIII fragment hybridizing to the probe (Figure 3.2.6), are indicated.

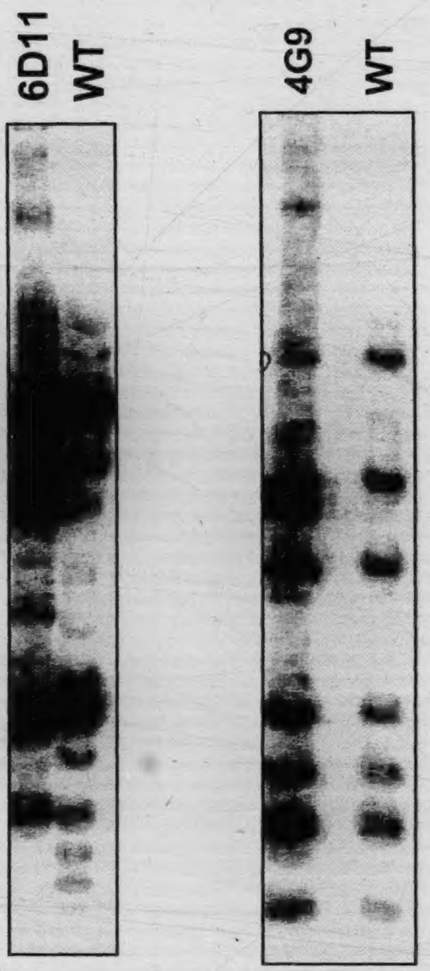


Figure 3.4.2 Southern analysis of the fosmids 4G9 and 6D11 identified in the pSMART-FOS genomic library of *P. polymyxa* PKB1. The fosmid DNA, and genomic DNA from wild-type PKB1 (WT) were digested with *Eco*RI before being subjected to Southern analysis. The membrane shown on the left hand side was probed using the *Eco*RI-digestion mixture of 6D11, while the membrane on the right hand side was hybridized with the gel-purified *Eco*RI fragments obtained from 4G9.

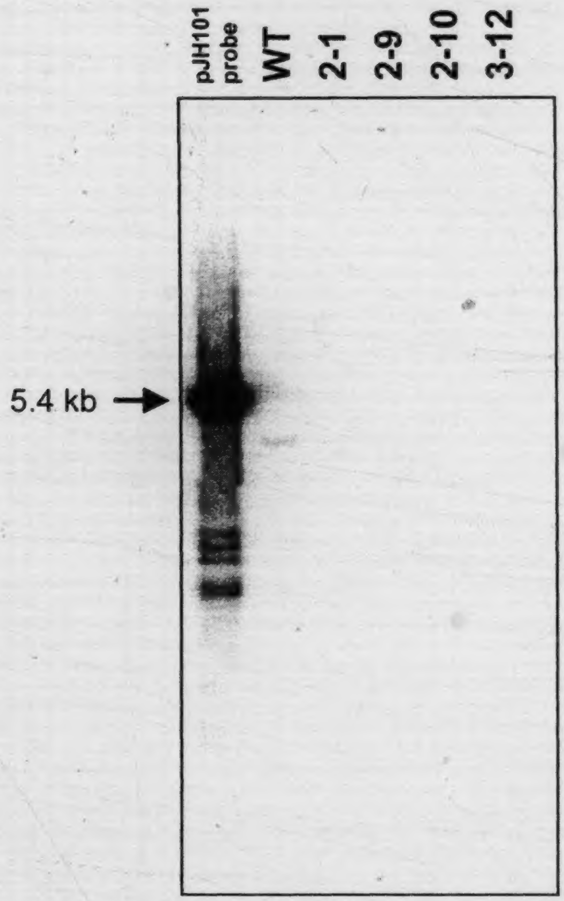


Figure 3.5.1 Southern analysis of the putative *fus-asn* single-crossover mutants. Genomic DNA from wild-type PKB1 (WT) and four Cm^R mutants (2-1, 2-9, 2-10, and 3-12) was digested with *Kpn*I before being subjected to Southern hybridization. The membrane was probed using the *Eco*RI-digested plasmid pJH101. The 5.4-kb hybridizing band indicated in the most left lane is the linearized form of pJH101.

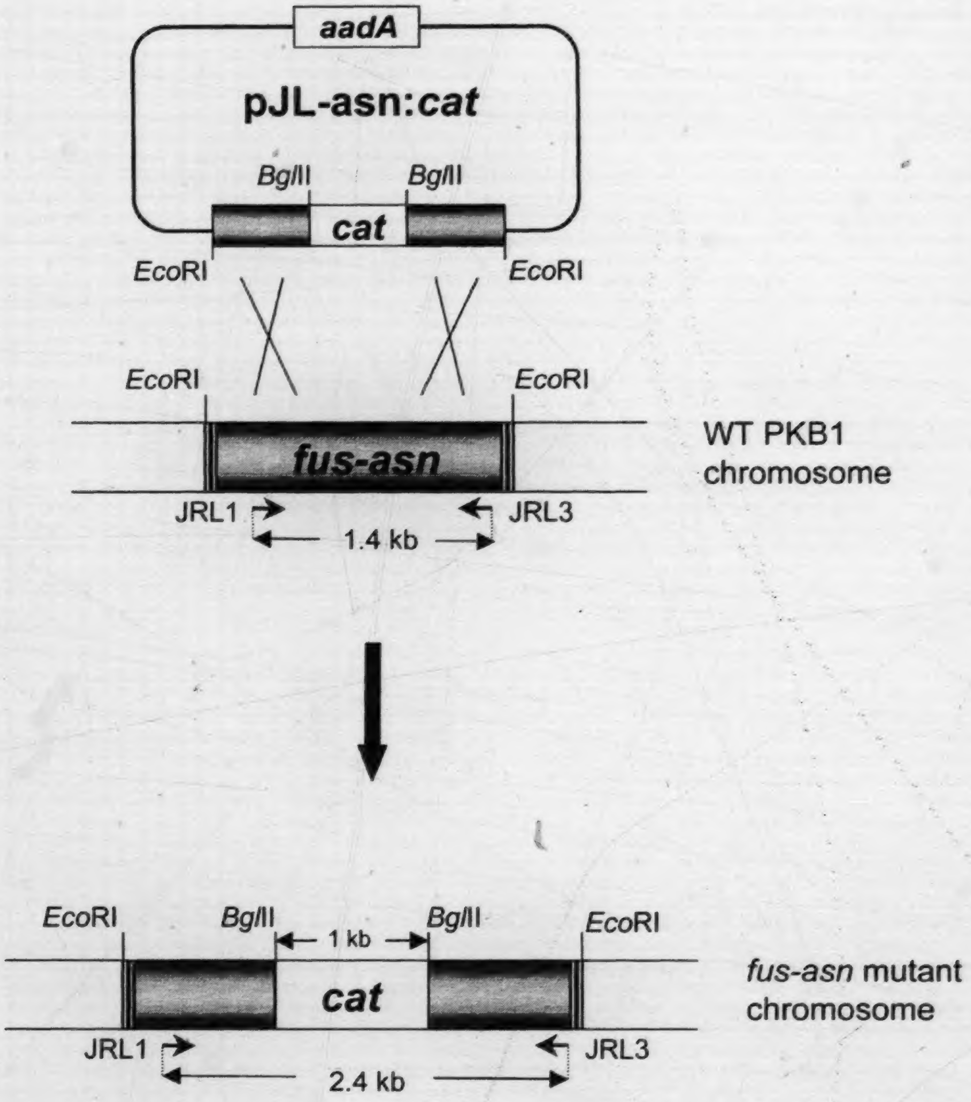

aadA
pJL-asn:*cat*
*Bgl*II
*Bgl*II
cat
*Eco*RI
*Eco*RI
*Eco*RI
*Eco*RI
fus-asn
WT PKB1 chromosome
JRL1
JRL3
1.4 kb
*Eco*RI
*Bgl*II
1 kb
*Bgl*II
*Eco*RI
cat
fus-asn mutant chromosome
JRL1
JRL3
2.4 kb

Figure 3.5.2 Gene disruption of the *fus-asn* fragment in *P. polymyxa* PKB1 via double crossover.

Figure 3.5.2 (continued) The plasmid pJL-asn:*cat* in which *fus-asn* is disrupted by insertion of the chloramphenicol resistance gene (*cat*) was transformed into the wild-type PKB1 strain. Replacement of the chromosomal copy of *fus-asn* by the disrupted copy via two crossovers should give rise to chloramphenicol resistant, but streptomycin sensitive *P. polymyxa* mutants. The *fus-asn* gene fragment is represented by the gray bar truncated on both sides by black double lines, and the surrounding chromosomal regions are represented by double lines. The locations of PCR primers used in subsequent PCR verification and the expected sizes of PCR products are also shown. The *aadA* gene is a streptomycin resistance marker taken from the plasmid pSETΩ.

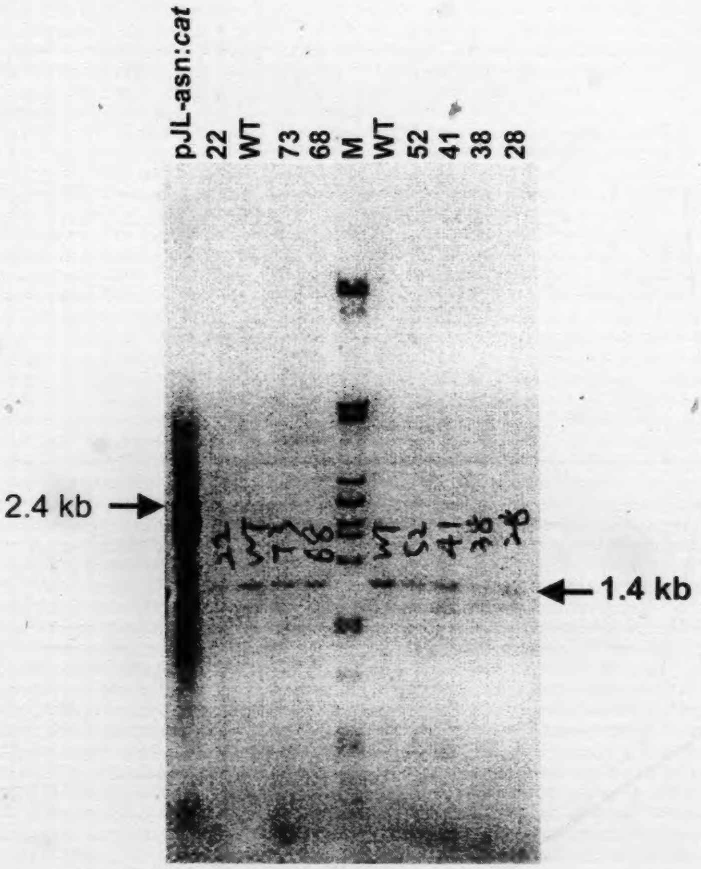


Figure 3.5.3 Analysis of the putative *fus-asn* double-crossover mutants by colony PCR. Cells harvested from fresh cultures of wild-type PKB1 (WT) and seven putative *fus-asn* mutants (22, 28, 38, 41, 52, 68, and 73) were lysed by boiling, and then supernatants of the cell lysates, as well as the plasmid pJL-asn:*cat*, were separately used as template in the PCR reaction with two *fus-asn*-specific primers JRL1 and JRL3 (Figure 3.5.2). Lane M: lambda DNA/*Pst*I marker.

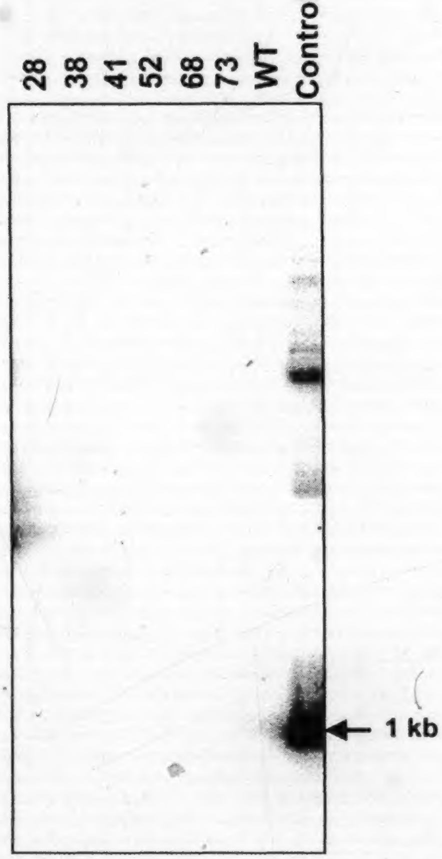


Figure 3.5.4 Southern analysis of the putative *fus-asn* double-crossover mutants. Genomic DNA from wild-type PKB1 (WT) and six Cm^R Str^S mutants (28, 38, 41, 52, 68, and 73), as well as the plasmid pJL-asn:*cat* (control), was digested with *Bgl*II. The fractionated and blotted DNA was probed using a *cat*-specific probe. The 1 kb hybridizing band indicated in the control lane is the *cat*-containing *Bgl*II fragment excised from the plasmid pJL-asn:*cat* (Figure 3.5.2).

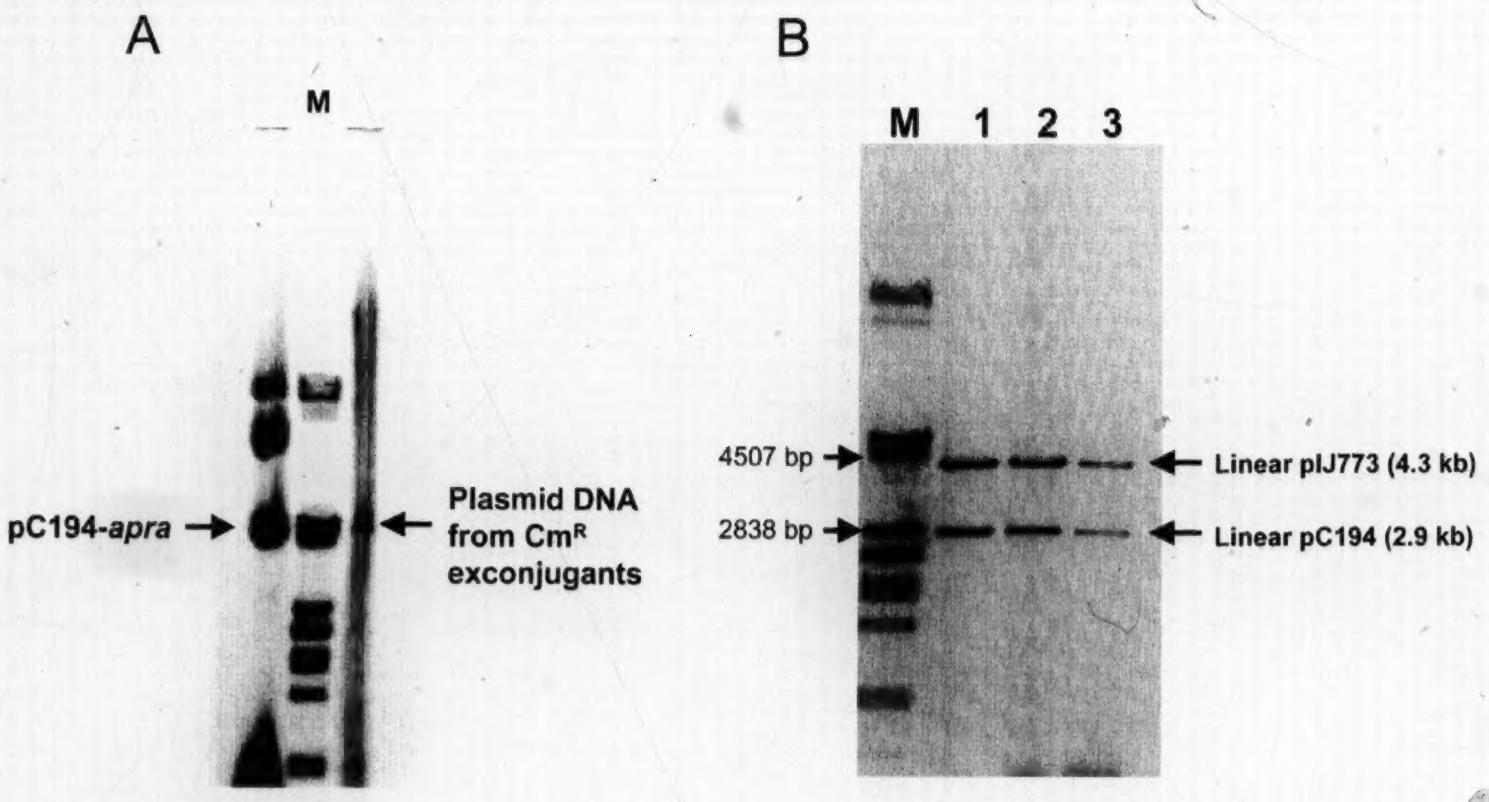
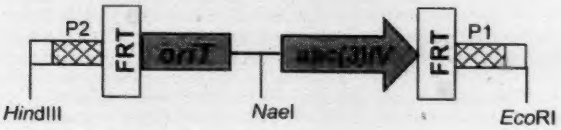

A
M
pC194-apra
Plasmid DNA from Cm^R exconjugants
B
M 1 2 3
4507 bp
2838 bp
Linear pIJ773 (4.3 kb)
Linear pC194 (2.9 kb)

Figure 3.5.5 Restriction analysis of the plasmid DNA isolated from chloramphenicol resistant (Cm^R) *P. polymyxa* exconjugants.

Figure 3.5.5 (continued) (A) Agarose gel electrophoresis of the uncut plasmid DNA isolated from *P. polymyxa* Cm^R exconjugants (right hand lane), resulting from conjugation of the *E. coli-Paenibacillus* shuttle vector pC194-*apra* into wild-type PKB1, and the uncut plasmid pC194-*apra* isolated from *E. coli* (left hand lane), which was constructed by fusion of the *Staphylococcus* plasmid pC194 (Cm^R) to the *E. coli* plasmid pIJ773 (Apra^R) at the *Hin*dIII site. *P. polymyxa* genomic DNA co-purified with the plasmid was observed as background smear. (B) The plasmid DNA obtained from Cm^R exconjugants was transformed into *E. coli* by electroporation. From the resultant apramycin and chloramphenicol resistant transformants, the plasmid DNA was purified and digested with *Hin*dIII (lanes 1&2), and then was electrophoresed on the agarose gel together with the *Hin*dIII-digested plasmid pC194-*apra* (lane 3). Lane M: lambda DNA/*Pst*I marker.

The original Apra^{R}*oriT* disruption cassette used in the *Streptomyces* system :


P2
FRT
oriT
aac(3)IV
FRT
P1
HindIII
NaeI
EcoRI

The hybrid Apra^{R}Cm^{R}*oriT* disruption cassette used in this study:

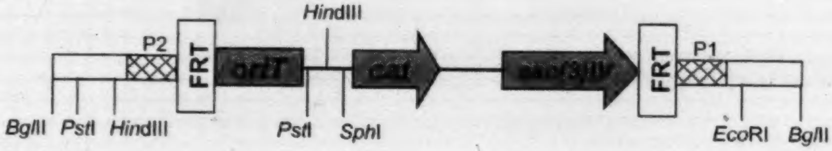

HindIII
P2
FRT
oriT
cat
aac(3)IV
FRT
P1
BglII
PstI
HindIII
PstI
SphI
EcoRI
BglII

Figure 3.5.6 Diagrammatic representation of the Apra^{R}*oriT* and Apra^{R}Cm^{R}*oriT* disruption cassettes. The Apra^{R}*oriT* disruption cassette was isolated from the plasmid pIJ773 as a 1.4-kb *Eco*RI/*Hin*dIII fragment, and the hybrid Apra^{R}Cm^{R}*oriT* disruption cassette was isolated from the plasmid pJL9 as a 2.5-kb *Bgl*II fragment. The hatched bars represent the priming sequences P1 (19 nt) and P2 (20 nt) that are supplied as the 3' ends of PCR primers used for amplification of the disruption cassette. FRT represents the FLP-recombinase recognition target site.

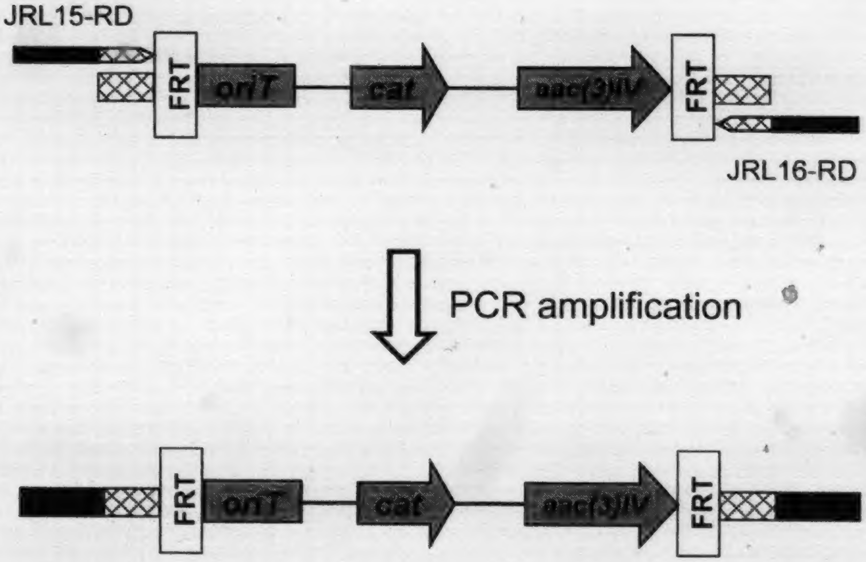

JRL15-RD
FRT
oriT
cat
aac(3)IV
FRT
JRL16-RD
PCR amplification
FRT
oriT
cat
aac(3)IV
FRT

Figure 3.5.7 PCR amplification of the Apra^{R}Cm^R*oriT* gene disruption cassette for preparation of *orf7* mutants using a modified Redirect© technology. By using a gel-purified 2.5-kb *Bgl*II fragment from the plasmid pJL9 as template, the Apra^{R}Cm^R*oriT* disruption cassette containing *acc(3)IV* (Apra^R), *cat* (Cm^R), *oriT*$_{RK2}$, and two FRT sequences was amplified with PCR primers JRL15-RD and JRL16-RD. Each primer includes a 39-nt extension sequence at the 5' end (black bar) identical to the region flanking the target DNA fragment, and a 19- or 20-nt priming sequence at the 3' end (hatched bar) identical to the disruption cassette.

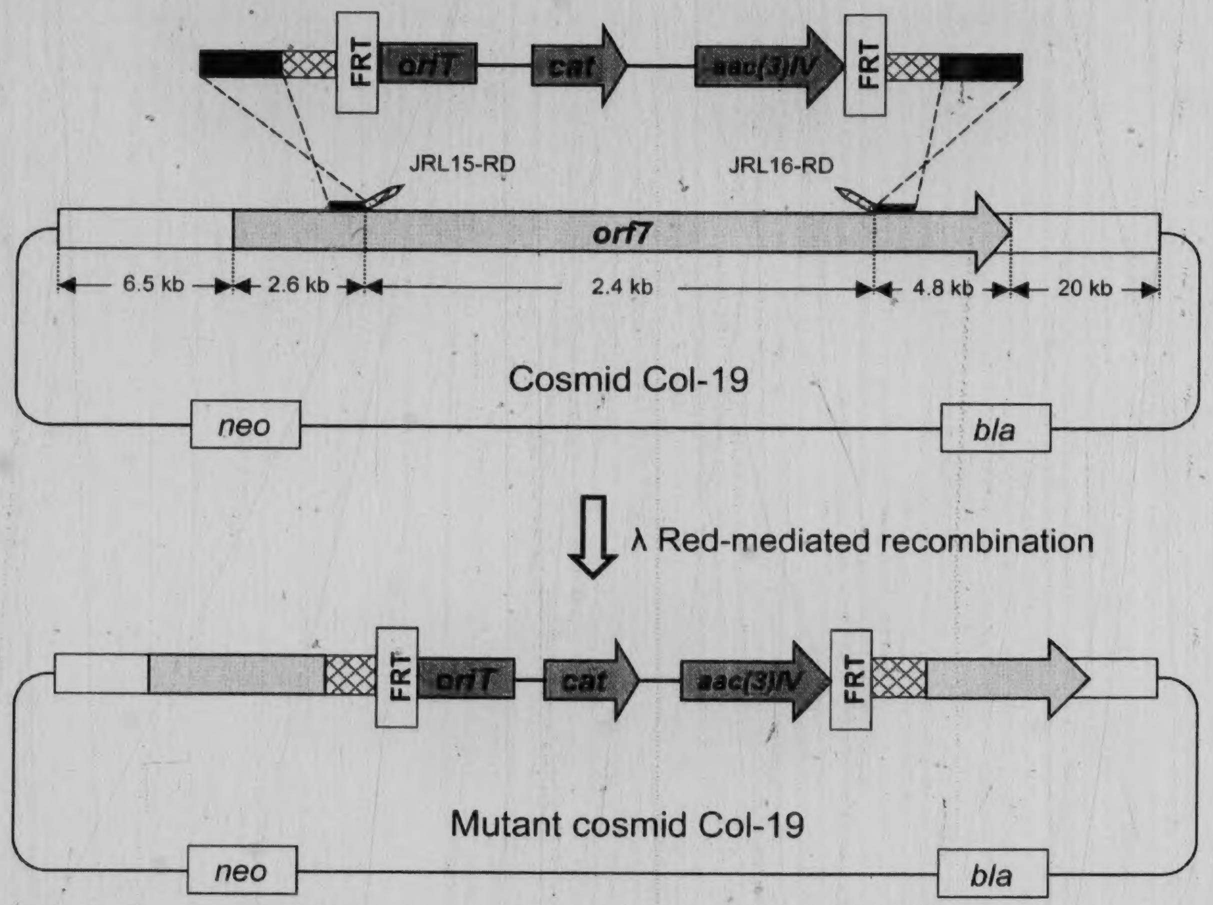

FRT
oriT
cat
aac(3)IV
FRT
JRL15-RD
JRL16-RD
orf7
6.5 kb
2.6 kb
2.4 kb
4.8 kb
20 kb
Cosmid Col-19
neo
bla
λ Red-mediated recombination
FRT
oriT
cat
aac(3)IV
FRT
Mutant cosmid Col-19
neo
bla

Figure 3.5.8 Construction of the mutant cosmid Col-19 containing a disrupted copy of *orf7*.

Figure 3.5.8 (continued) A 2.4-kb segment internal to *orf7* (grey arrow) was targeted for PCR primer design (JRL-15RD and JRL-16RD). The PCR-amplified Apra^{R}Cm^R*oriT* disruption cassette was transformed into *E. coli* BW25113/pIJ790 carrying the cosmid Col-19, in which the λ Red functions encoded by pIJ790 promoted homologous recombination between the linear PCR product and the cosmid Col-19, resulting in replacement of the target segment internal to *orf7* by the disruption cassette. The white bars represent the rest of the cosmid insert DNA. The antibiotic resistance genes (*neo* for Kan^R, *bla* for Amp^R) on the cosmid vector are also shown. The figure is not drawn to scale.

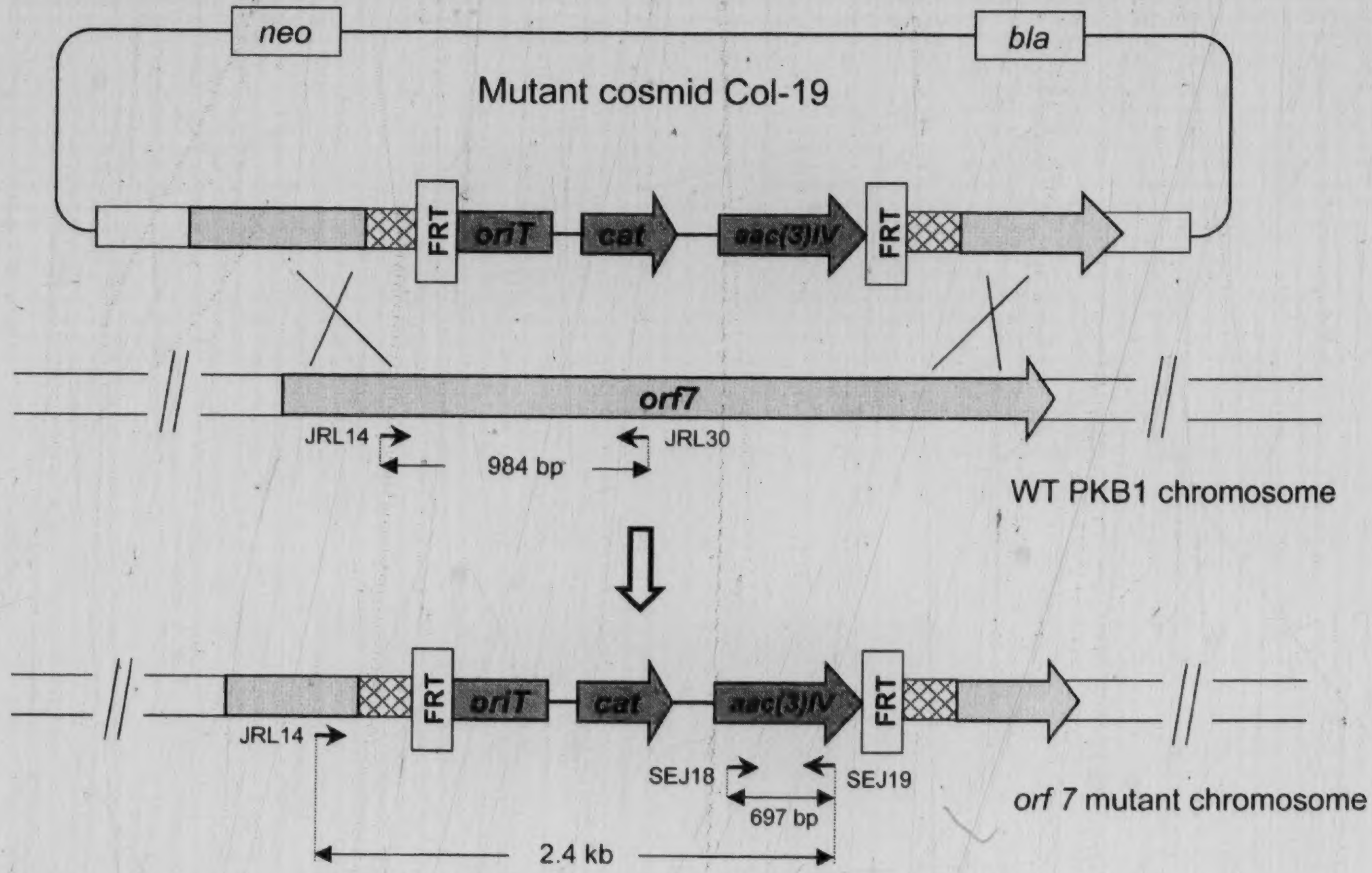

neo
bla
Mutant cosmid Col-19
FRT
oriT
cat
aac(3)IV
FRT
orf7
JRL14
JRL30
984 bp
WT PKB1 chromosome
FRT
oriT
cat
aac(3)IV
FRT
JRL14
SEJ18
SEJ19
697 bp
2.4 kb
orf 7 mutant chromosome

Figure 3.5.9 Gene disruption of the chromosomal copy of *orf7* in *P. polymyxa* PKB1.

Figure 3.5.9 (continued) The mutant cosmid Col-19 carrying the disrupted-copy of *orf7* was mobilized into PKB1 via intergeneric conjugation, where exchange with the wild-type copy occurred. Cm^R *P. polymyxa* exconjugants were selected, and successful integration of the disruption cassette in the chromosome was confirmed by PCRs and Southern hybridization. The chromosomal DNA is indicated by double lines. The locations of PCR primers and the expected sizes of PCR products were shown.

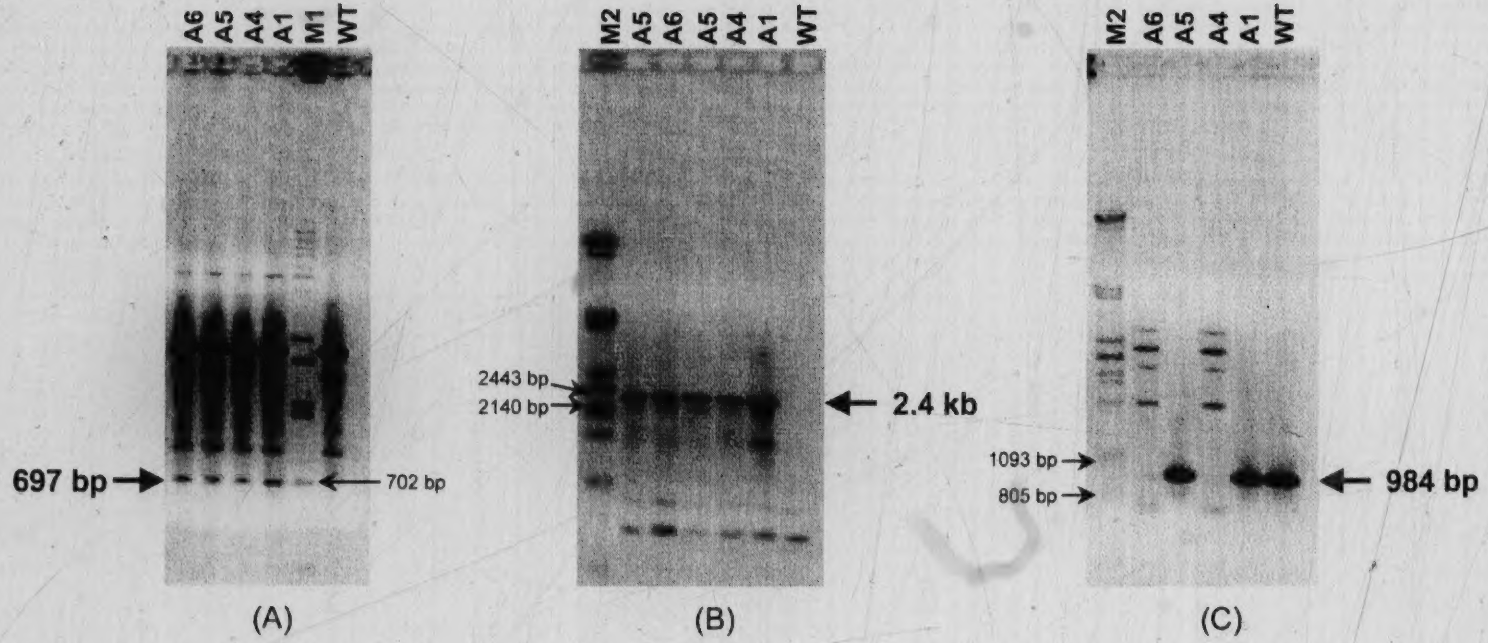

A6
A5
A4
A1
M1
WT
697 bp
702 bp
(A)
M2
A5
A6
A5
A4
A1
WT
2443 bp
2140 bp
2.4 kb
(B)
M2
A6
A5
A4
A1
WT
1093 bp
805 bp
984 bp
(C)

Figure 3.5.10 PCR verification of the *P. polymyxa orf7* mutants.

Figure 3.5.10 (continued) Genomic DNA from four independent mutants A1, A4, A5, and A6, as well as the wild-type PKB1 strain (WT), was amplified by using a pair of *aac(3)IV*–specific primers SEJ18 and SEJ19 (A), primer SEJ19 and a primer JRL14 flanking the segment targeted for gene replacement (B), primer JRL14 and an *orf7*-specific primer JRL30 (C), respectively. The expected sizes of PCR products are indicated in all three reactions. See Figure 3.5.9 for the locations of PCR primers. Lane M1: lambda DNA/*Bst*EII marker; lane M2: lambda DNA/*Pst*I marker.

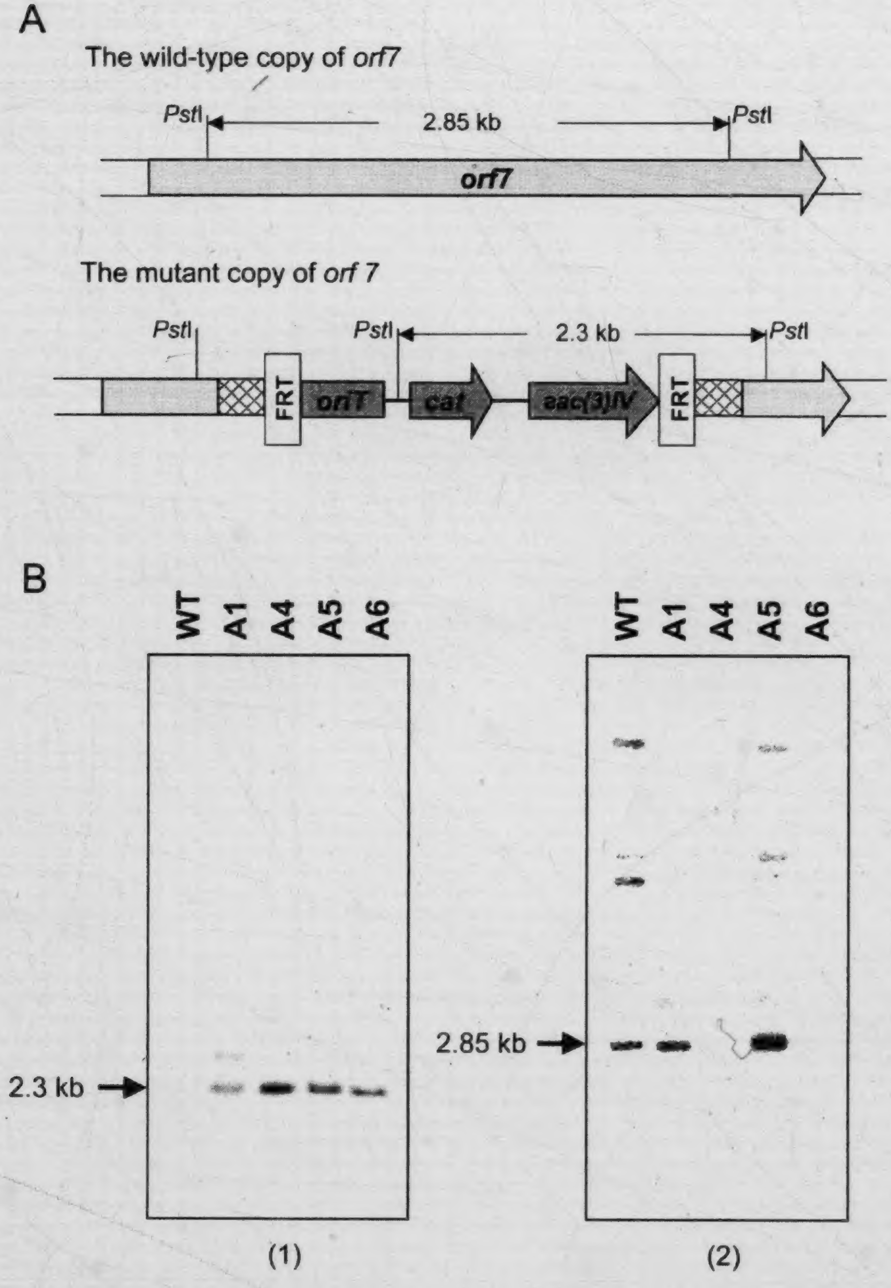

A
The wild-type copy of orf7
PstI
2.85 kb
PstI
orf7
The mutant copy of orf 7
PstI
PstI
2.3 kb
PstI
FRT
oriT
cat
aac(3)IV
FRT
B
WT
A1
A4
A5
A6
2.3 kb
(1)
WT
A1
A4
A5
A6
2.85 kb
(2)

Figure 3.5.11 Southern analysis of the *P. polymyxa orf7* mutants.

Figure 3.5.11 (continued) (A) Diagrammatic representation of the chromosomal *orf7* region in the wild-type PKB1 strain and the *orf7* mutant. The double lines represent the rest of the chromosomal DNA. (B) Southern hybridization results of the *Pst*I digested genomic DNA from the wild-type PKB1 strain (WT) and the four *orf7* mutants (A1, A4, A5, and A6). The membrane with bound DNA was first probed with a *cat*-specific probe (B-1), and then stripped and re-probed with a 2.85-kb *Pst*I fragment internal to *orf7* (B-2).

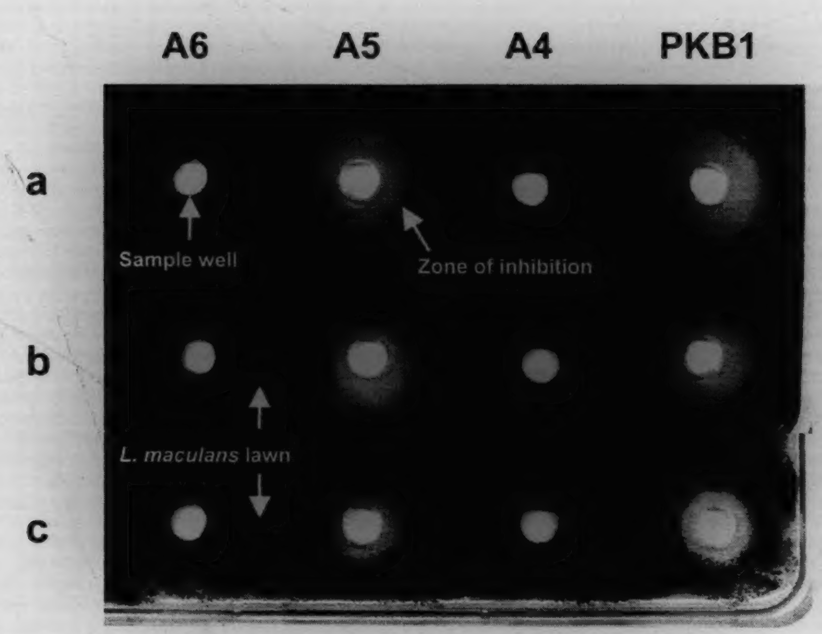

A6
A5
A4
PKB1
a
b
c
Sample well
Zone of inhibition
L. maculans lawn

Figure 3.5.12 The antifungal activity of A4, A5, and A6 mutants carrying the *orf7* disruption compared to the activity of PKB1 strain. Both mutants and the wild-type strain were grown in PDB-soy medium for 72 h, and then a well bioassay was used to evaluate the antifungal activity of the bacterial culture (a), the cell and spore pellet (b), and the culture supernatant (c) against *L. maculans*. The zones of inhibition were observed after 3 days.

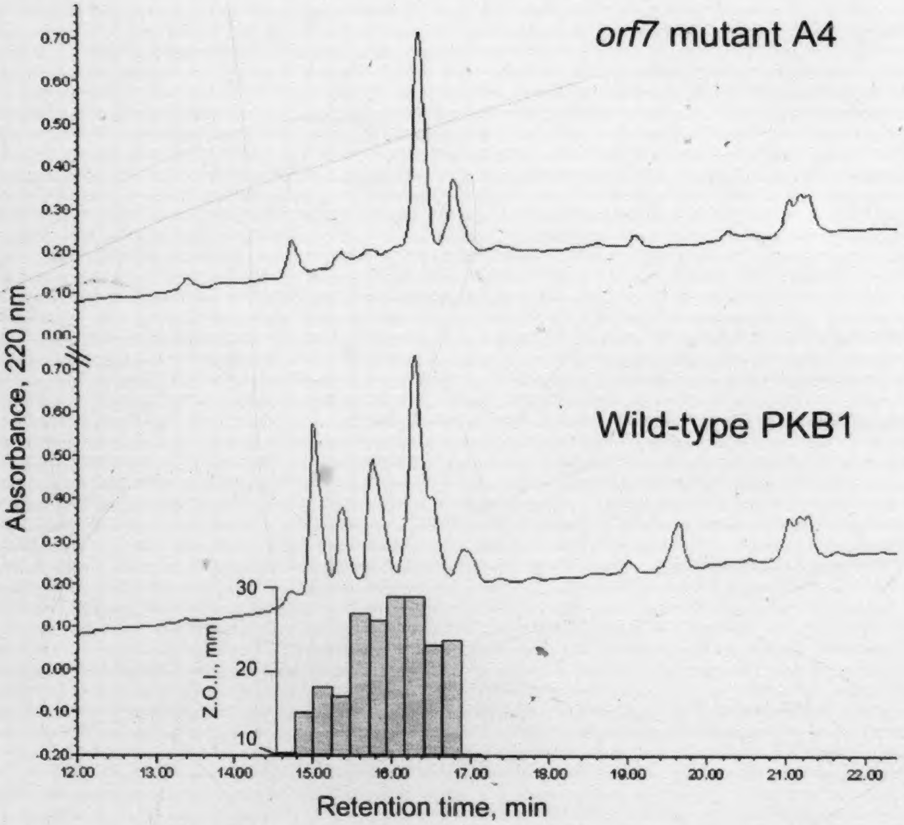


Figure 3.5.13 HPLC analysis of methanol extracts from the wild-type and A4 mutant cultures. Methanol extracts of whole PDB-soy medium-grown cultures were partially purified before analysis by absorption to a Sep-Pak C_{18} cartridge. Concentrated bioactive material was then separated by reversed-phase HPLC with gradient elution. Fractions were collected across the elution profile and bioassayed against *L. maculans*. The bioassay sample wells were 10 mm in diameter, which was the minimum zone of inhibition (Z.O.I) observable.

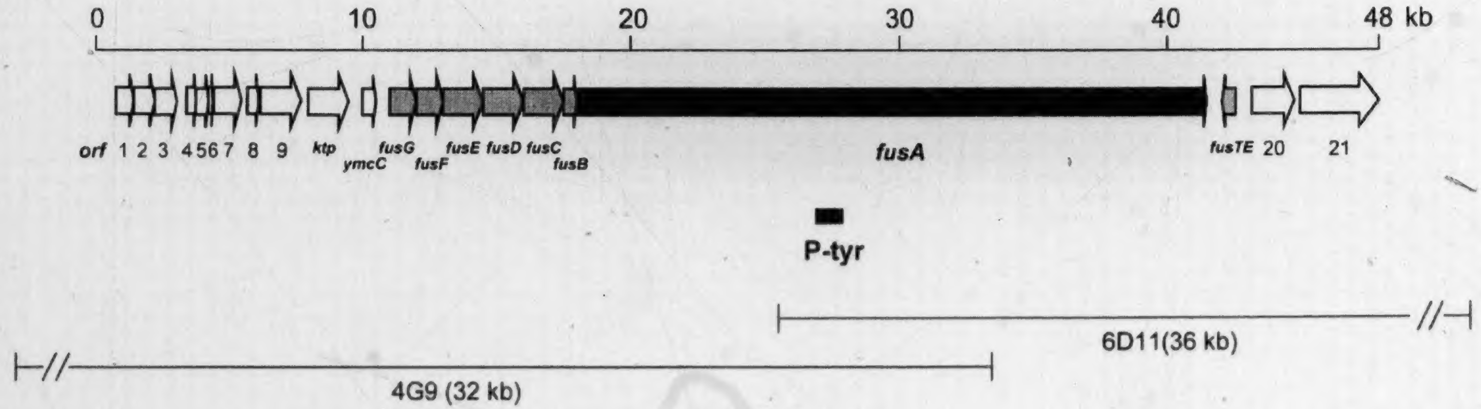

0
10
20
30
40
48 kb
orf 1 2 3 4567 8 9
ktp
ymcC
fusG
fusF
fusE fusD fusC
fusB
fusA
fusTE 20
21
P-tyr
6D11(36 kb)
4G9 (32 kb)

Figure 3.6.1 ORF arrangement of the fusaricidin biosynthetic gene cluster. The black arrow represents the peptide synthetase gene *fusA* whereas other fusaricidin biosynthetic genes are shown in gray and flanking regions are in white. The P-tyr probe used for screening the pSMART-FOS genomic library of *P. polymyxa* PKB1, and the overlapping fosmid inserts isolated in this study are shown. See Table 3.5 for proposed functions of assigned ORFs. The figure is not drawn to scale.

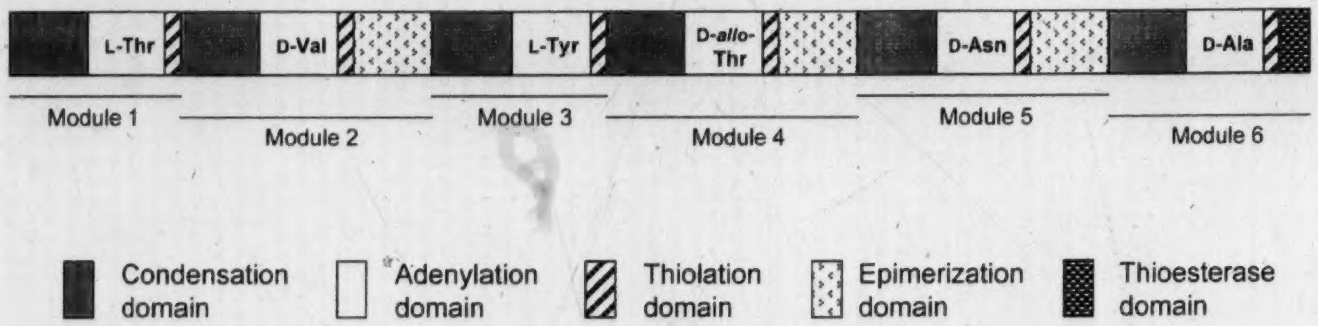
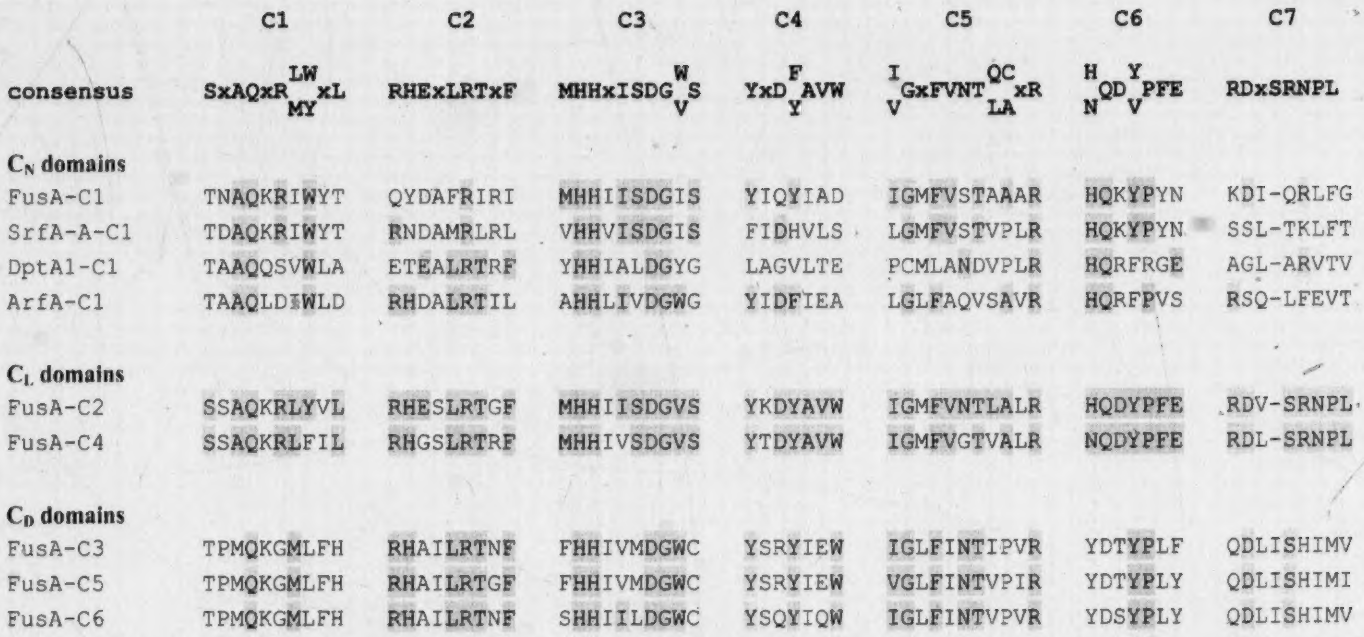

L-Thr
D-Val
L-Tyr
D-allo-Thr
D-Asn
D-Ala
Module 1
Module 2
Module 3
Module 4
Module 5
Module 6
Condensation domain
Adenylation domain
Thiolation domain
Epimerization domain
Thioesterase domain

Figure 3.6.2 Module and domain organization of the fusaricidin synthetase encoded by *fusA*.



	C1	C2	C3	C4	C5	C6	C7
consensus	$\mathrm{SxAQxR}^{\mathrm{LW}}_{\mathrm{MY}}\mathrm{xL}$	**RHExLRTxF**	$\mathrm{MHHxISDG}^{\mathrm{W}}_{\mathrm{V}}\mathrm{S}$	$\mathrm{YxD}^{\mathrm{F}}_{\mathrm{Y}}\mathrm{AVW}$	${}^{\mathrm{I}}_{\mathrm{V}}\mathrm{GxFVNT}^{\mathrm{QC}}_{\mathrm{LA}}\mathrm{xR}$	${}^{\mathrm{H}}_{\mathrm{N}}\mathrm{QD}^{\mathrm{Y}}_{\mathrm{V}}\mathrm{PFE}$	**RDxSRNPL**
C$_N$ domains							
FusA-C1	TNAQKRIWYT	QYDAFRIRI	MHHIISDGIS	YIQYIAD	IGMFVSTAAAR	HQKYPYN	KDI-QRLFG
SrfA-A-C1	TDAQKRIWYT	RNDAMRLRL	VHHVISDGIS	FIDHVLS	LGMFVSTVPLR	HQKYPYN	SSL-TKLFT
DptA1-C1	TAAQQSVWLA	ETEALRTRF	YHHIALDGYG	LAGVLTE	PCMLANDVPLR	HQRFRGE	AGL-ARVTV
ArfA-C1	TAAQLDSWLD	RHDALRTIL	AHHLIVDGWG	YIDFIEA	LGLFAQVSAVR	HQRFPVS	RSQ-LFEVT
C$_L$ domains							
FusA-C2	SSAQKRLYVL	RHESLRTGF	MHHIISDGVS	YKDYAVW	IGMFVNTLALR	HQDYPFE	RDV-SRNPL
FusA-C4	SSAQKRLFIL	RHGSLRTRF	MHHIVSDGVS	YTDYAVW	IGMFVGTVALR	NQDYPFE	RDL-SRNPL
C$_D$ domains							
FusA-C3	TPMQKGMLFH	RHAILRTNF	FHHIVMDGWC	YSRYIEW	IGLFINTIPVR	YDTYPLF	QDLISHIMV
FusA-C5	TPMQKGMLFH	RHAILRTGF	FHHIVMDGWC	YSRYIEW	VGLFINTVPIR	YDTYPLY	QDLISHIMI
FusA-C6	TPMQKGMLFH	RHAILRTNF	SHHIILDGWC	YSQYIQW	IGLFINTVPVR	YDSYPLY	QDLISHIMV

Figure 3.6.3 Comparison of the conserved core motifs within the condensation (C) domains of FusA. Alternative amino acids at a particular position in the core motifs are indicated. The N-terminal C domains of the surfactin synthetase SrfA-A from *Bacillus subtilis* (GenBank accession no. BAA02522), the daptomycin synthetase DptA from *Streptomyces roseosporus* (GenBank accession no. AAX31557), and the arthrofactin synthetase ArfA from *Pseudomonas* sp. MIS38 (GenBank accession no. BAC67534), were used in comparison with the initial C domain of FusA. C$_L$ and C$_D$ domains of FusA represent condensation domains following modules incorporating L- and D-amino acid residues, respectively. Shading indicates sequence identity with the consensus sequence. The condensation domains in each peptide synthetase are numbered according to their order in the protein.

FusA-T1	YLAPRTAVEAQMVLIWQDILGVARVGVRDNFFEIGG**H**SLRATVLVSRIHKELGCSISLREVFQSPTVESLAQLVKKHIPTLYE
FusA-T3	SEAPRTPVEIQLAEIWQEVLGVESAGVKDNFFHFGG**H**SLRAALLVSRIRKEMNREISLRAVFESPTIEGLARAIEGYTPLNFE
FusA-T6	YIPPRTQTEVELAQIWTEVLGVQEIGVKDHFFELGG**H**SLKVLGLIQKISSGMGVQLPIQLVFNLPTVEEMAHEISKLRAKDAP
FusA-T2	YTAPRTPAEEALASVWQGVLGAKQVGIHDNFFELGG**D**SIKAIQVSSRLLQAGYR-LEMKQLFKSPTIAELGAEIQTAVHMAEQ
FusA-T4	YEPPRTKAEEALASVWQGVLGAQQVGIHDHFFDLGG**D**SIKAIQVSSRLFQAGYK-LEMHDLFKYPTLAELSPYLQAAGRTAEQ
FusA-T5	FTAPRTDVENILASIWQGVLGVPLVGIHDNFFELGG**D**SIKSLQVSSRLLQAGYK-IEMKDLFGYPTIAELAQRVSVVSRIADQ

Figure 3.6.4 Alignment of the amino acid sequences of six thiolation (T) domains in FusA. Consensus core motifs are underlined. The residues in front of the conserved serine residues are shown in bold. Shading indicates sequence differences in the downstream regions of the core motifs between these two groups of T domains.

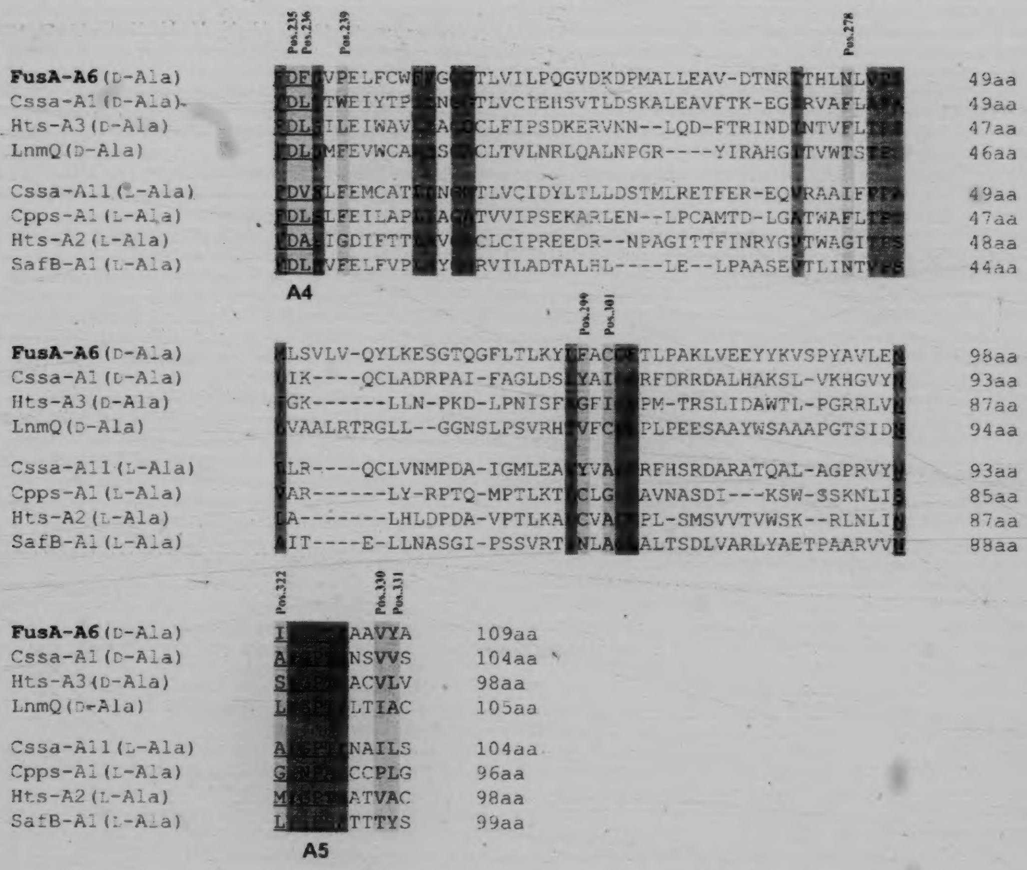

Pos.235
Pos.236
Pos.239
Pos.278
Pos.299
Pos.301
Pos.322
Pos.330
Pos.331
FusA-A6 (D-Ala)
Cssa-A1 (D-Ala)
Hts-A3 (D-Ala)
LnmQ (D-Ala)
Cssa-A11 (L-Ala)
Cpps-A1 (L-Ala)
Hts-A2 (L-Ala)
SafB-A1 (L-Ala)
A4
A5
49aa
49aa
47aa
46aa
49aa
47aa
48aa
44aa
98aa
93aa
87aa
94aa
93aa
85aa
87aa
88aa
109aa
104aa
98aa
105aa
104aa
96aa
98aa
99aa

Figure 3.6.5 Alignment of the amino acid sequences in core motifs A4 and A5 of FusA-A6 compared to other known A domains assigned to activate L- or D-Ala. Deduced specificity-conferring residues are shaded in light grey, and structural anchor residues are shaded in dark grey. Conserved motif sequences A4 and A5 are underlined. Abbreviations for protein names are as follows: Cssa, cyclosporine synthetase from *Tolypocladium inflatum* (GenBank accession no. CAA82227); Hts, HC-toxin synthetase from *Cochliobolus carbonum* (GenBank accession no. AAA33023); Cpps, d-lysergyl peptide synthetase from *Claviceps purpurea* (GenBank accession no. CAB39315); SafB, saframycin Mx1 synthetase B from *Myxococcus xanthus* (GenBank accession no. AAC44128). LnmQ, the stand-alone adenylation domain of leinamycin synthetase from *Streptomyces atroolivaceus* (GenBank accession no. AAN85530). The A domains in each peptide synthetase are numbered according to their order in the protein.

SVAVQPTI**RIDELELITPEEKSQIIEVWGDTAAPYPREQTLHGIF** 45aa

JRL38

EEKAALTPDRTALIYGETELTYGELHQQANRLARTLRAQGVRPDQ 90aa

A1

PVGIMVERSLEMIIGIHAILKAGGAYVPIDPEFPEDRIRHMLEDS 135aa

A2

GAKLLLTKNHLKDRFPFTGTILALDDPQMYHADSSNLEPIAGPEH 180aa

LAYIIYTSGSTGKPKGVMIEHRSAVHTLSQLEAEYPMLAGDRFLL 225aa

A3

KTTFTFDFSVPELFCWFFGQGTLVILPQGVDKDPMALLEAVDTNR 270aa

A4

ITHLNLVPSMLSVLVQYLKESGTQGFLTLKYLFACGETLPAKLVE 315aa

EYYKVSPYAVLENIYGPTEAAVYATRYTTSLETAALTHVPIGKPY 360aa

A5

ANVQVWMMDSASQVSPVGVPGELCIAGEGVARGYFNQPDLTAEKF 405aa

A6

IPHPYKPGARIYRTGDLARWLPDGNIEYLGRIDHQVKIRGYRIEL 450aa

A7 A8

GEVEAQILKVPSVQEAVALALADSTGSTQLCAYFVAEEGLAAGVL 495aa

REALASELPSYMIPTAFVQLAQMPLNPNGKLDRKALPAPEALLRS 540aa

A9 A10

TAEYIPPRTQTEVELAQIWTEVLGVQEIGVKDHFFELGGHSLKVL 585aa

JRL37 T

GLIQKISSGMGVQLPLQLVFNLPTVEEMAHEISKLRAKDAP 626aa

Figure 3.6.6 Amino acid sequence of the adenylation (A) and thiolation (T) domains in the sixth module (FusA-M6) of fusaricidin synthetase FusA. Conserved core motifs of the A and T domains are underlined. The locations of two PCR primers, JRL37 and JRL38 used to amplify a DNA fragment corresponding to the A domain of FusA-M6, are indicated.

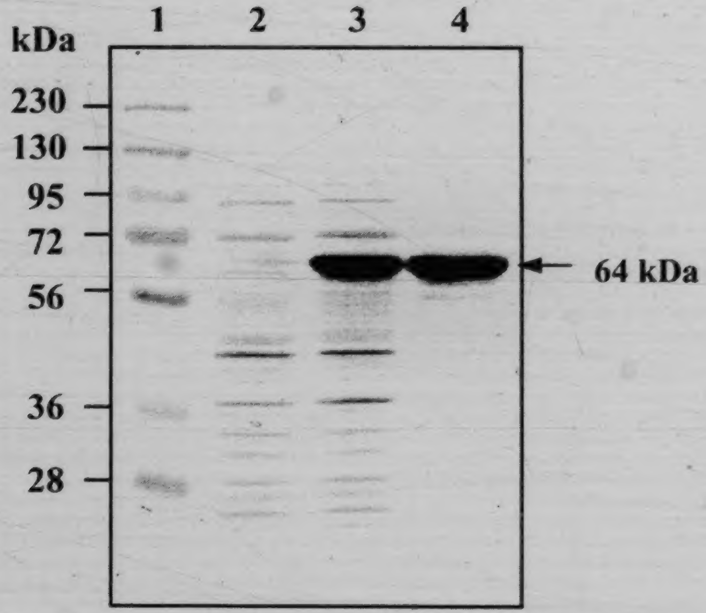


Figure 3.6.7 Overeproduction and purification of His-tagged recombinant FusA-A6 protein. Lane 1: prestained protein ladder; lane 2: whole cell extract obtained before IPTG induction; lane 3: whole cell extract obtained after 4 h-induction with IPTG; lane 4: protein purified by Ni^{2+}-affinity chromatography.

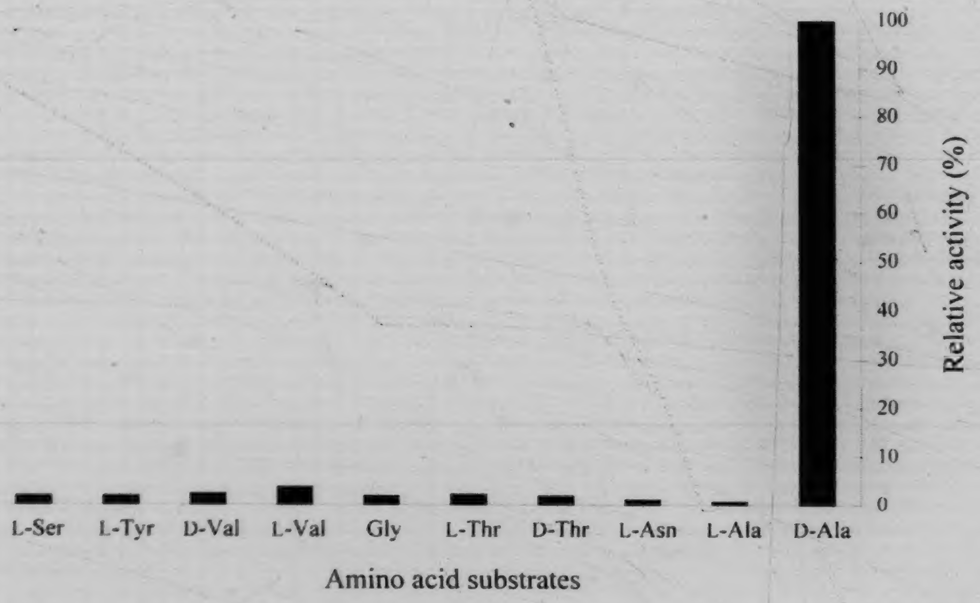


Figure 3.6.8 Relative activities of the purified FusA-A6 protein for various amino acid substrates determined using the ATP-PPi exchange assay. The control reaction without amino acids was subtracted from all reactions with amino acid substrates.

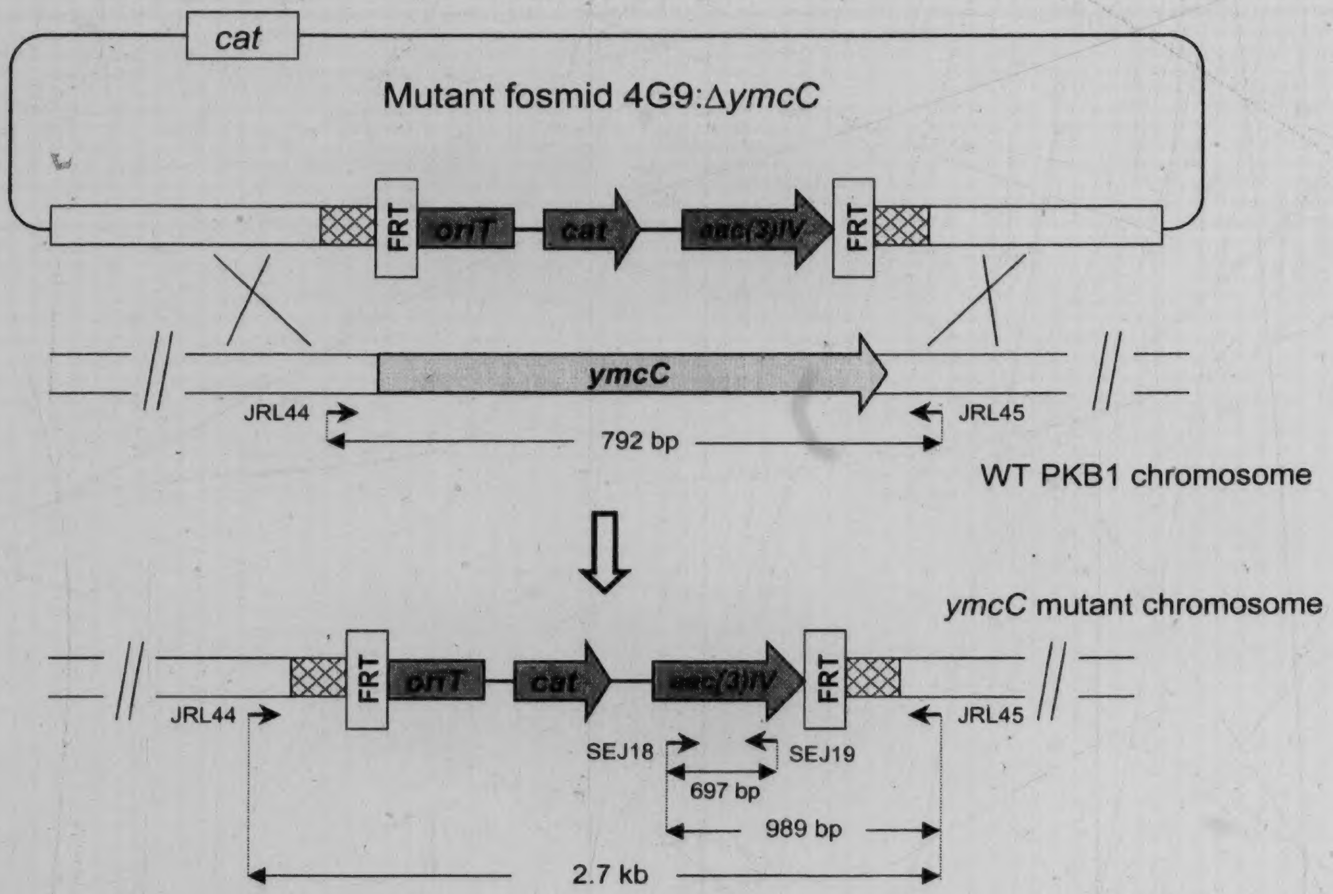

cat
Mutant fosmid 4G9:Δ*ymcC*
FRT
oriT
cat
aac(3)IV
FRT
ymcC
JRL44
JRL45
792 bp
WT PKB1 chromosome
ymcC mutant chromosome
FRT
oriT
cat
aac(3)IV
FRT
JRL44
JRL45
SEJ18
SEJ19
697 bp
989 bp
2.7 kb

Figure 3.6.9 PCR-targeted mutagenesis of *ymcC* in *P. polymyxa* PKB1.

Figure 3.6.9 (continued) The mutant fosmid 4G9:Δ*ymcC*, in which *ymcC* was deleted and replaced by the Apra^{R}Cm^R*oriT* disruption cassette was obtained using a PCR-targeting protocol (Figure 3.5.8), and then was introduced into the wild-type PKB1 strain via intergeneric conjugation. Homologous recombination between the mutant fosmid and the PKB1 chromosome resulted in replacement of the wild-type copy of *ymcC* by the disruption cassette, which was then verified by PCR reactions. The locations of PCR primers used and the expected sizes of PCR products were shown. The double lines represent the chromosomal region flanking *ymcC*.

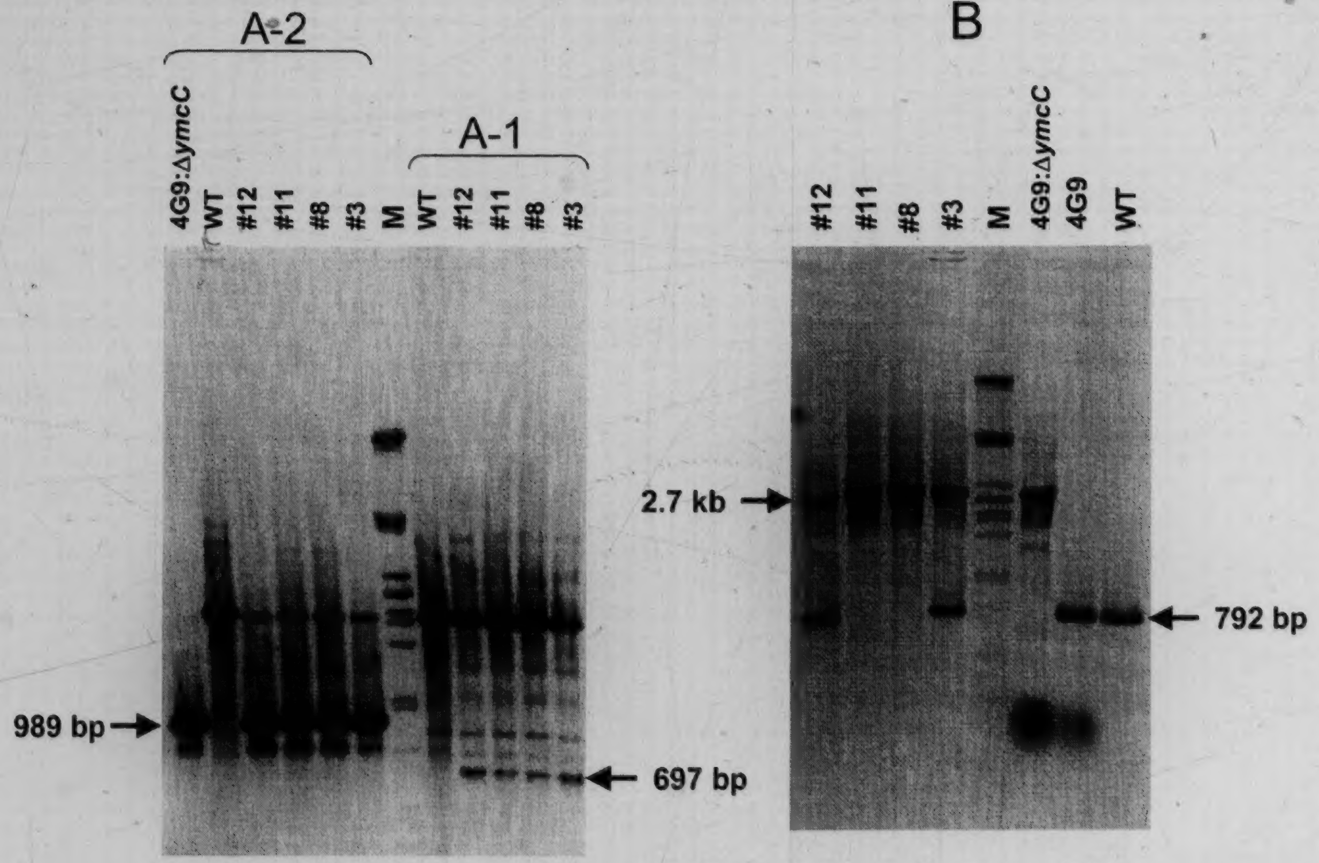

A-2
A-1
B
4G9:ΔymcC
WT
#12
#11
#8
#3
M
WT
#12
#11
#8
#3
#12
#11
#8
#3
M
4G9:ΔymcC
4G9
WT
989 bp
697 bp
2.7 kb
792 bp

Figure 3.6.10 Verification of the *P. polymyxa ymcC* mutants by PCR reactions.

Figure 3.6.10 (continued) Genomic DNA from the wild-type PKB1 strain (WT) and the four independent *ymcC* mutants (#3, #8, #11, #12), as well as the fosmid DNA from 4G9 and 4G9:Δ*ymcC*, was amplified by using a pair of *aac(3)IV*–specific primers SEJ18 and SEJ19 (A-1), primer SEJ18 and a primer flanking the locus of *ymcC*, JRL45 (A-2), primer JRL45 and another flanking locus-specific primer JRL44 (B), respectively. See Figure 3.6.9 for the locations of PCR primers. Lane M: lambda DNA/*Pst*I marker.

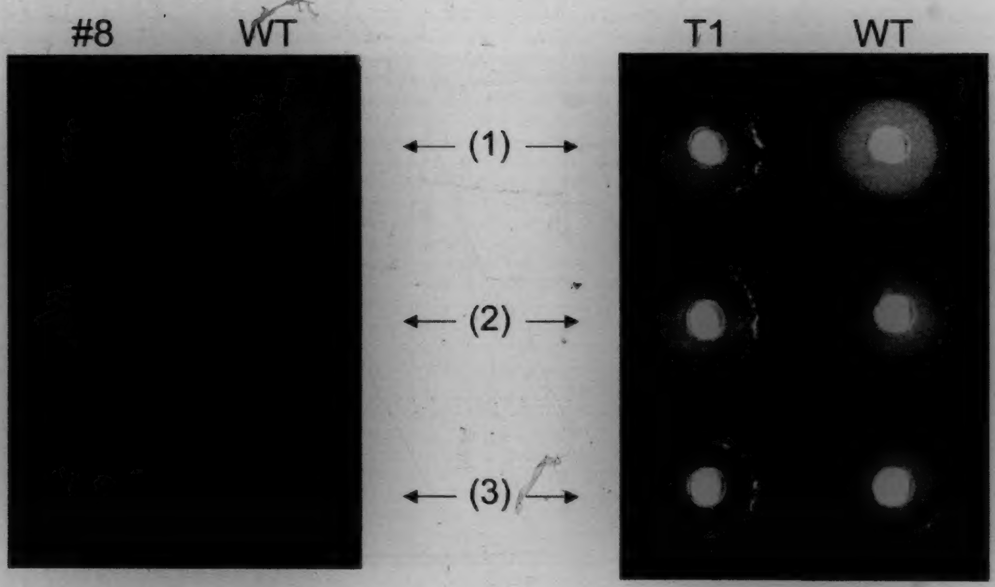


Figure 3.6.11 Comparison of the antifungal activity of wild-type PKB1 with the *ymcC* and *ktp* mutants. The *ymc*C mutant #8 and the *ktp* mutant T1 were obtained by gene replacement with the Apra^{R}Cm^R*oriT* disruption cassette using a PCR-targeting mutagenesis method. After grown in PDB-soy medium for 72 h, the wild-type and mutant cultures were bioassayed against *L. maculans*. Methanol extracts of the bacterial culture (1), the cell and spore pellet (2), and the culture supernatant (3) were used in the well bioassay, and the antifungal inhibition zones were observed after 3 days of incubation.

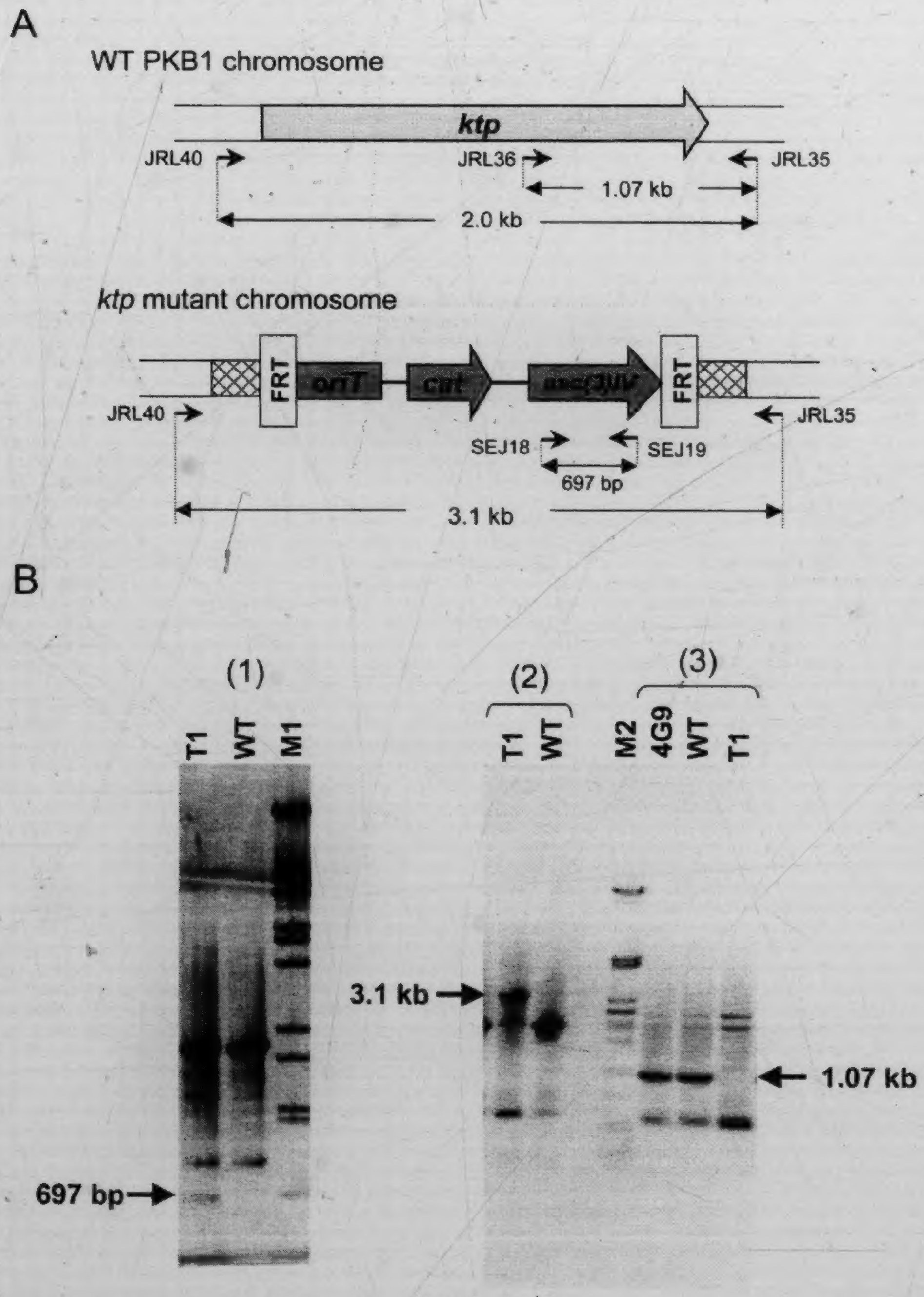

A
WT PKB1 chromosome
ktp
JRL40
JRL36
JRL35
1.07 kb
2.0 kb
ktp mutant chromosome
FRT
oriT
cat
aac(3)IV
FRT
JRL40
JRL35
SEJ18
SEJ19
697 bp
3.1 kb
B
(1)
(2)
(3)
T1
WT
M1
T1
WT
M2
4G9
WT
T1
3.1 kb
1.07 kb
697 bp

Figure 3.6.12 PCR analysis of the *P. polymyxa ktp* mutant.

Figure 3.6.12 (continued) (A) Diagrammatic representation of the region encompassing *ktp* (grey arrow) in the wild-type PKB1 chromosome, and the chromosomal region encompassing the Apra^{R}Cm^R*oriT* disruption cassette that had replaced the wild-type copy of *ktp* in the mutant. The double lines represent the rest of *P. polymyxa* chromosome. The locations of PCR primers used and the expected sizes of PCR products were indicated. (B) Genomic DNA from the wild-type PKB1 strain (WT) and the *ktp* mutant T1, as well as the fosmid 4G9, was amplified by using a pair of *aac(3)IV*–specific primers SEJ18 and SEJ19 (1), two primers flanking the locus of *ktp*, JRL35 and JRL40 (2), primer JRL35 and a *ktp*-specific primer JRL36 (3), respectively. Lane M1: lambda DNA/*Bst*EII marker; lane M2: lambda DNA/*Pst*I marker.

Table 3.1 Predicted substrate specificities of the adenylation (A) domains derived from the PCR-amplified peptide synthetase gene fragments, based on the selectivity-conferring codes of NRPSs.

A domain	Positions of amino acid residues involved in substrate recognition									Predicated substrate
	235	236	239	278	299	301	322	330	331	
B1J6-2	D	A	W	I	F	G	G	M	P	Glu
B1J6-3	D	A	W	I	F	G	A	I	T	Glu
B1J6-5	D	V	C	E	T	G	T	I	E	Orn
B1J6-7	D	F	W	N	*	G	M	*	*	Thr
B1J6-10	D	L	T	K	I	G	E	V	G	Asn
B1J6-14	D	V	G	E	I	G	A	P	*	Orn
B1J6-17	D	F	W	N	I	G	M	V	H	Thr
B1J6-19	D	V	G	E	I	G	S	I	D	Orn

" * " represents unidentified amino acids.

Table 3.2 Summary of the proposed functions of ORFs identified within the insert of cosmid Col-8.

ORF	**Nucleotide position (start-stop)**	**Top BLAST hit**	**Proposed function**
orf1	108-1037	Phosphotransferase system, glucose-specific IIABC component [*Bacillus thuringiensis serovar israelensis* ATCC 35646] EAO51304	Sugar transport
orf2	1489-4260	Mycosubtilin synthetase, MycB [*Bacillus subtilis*] AAF08796	Peptide synthetases
orf5	6633-9023	Mycosubtilin synthetase, MycC [*Bacillus subtilis*] AAF08797	
or6	9101-19921	Gramicidin S synthetase 2 [*Brevibacillus brevis*] BAA06146	Polymyxin biosynthesis?
orf7	20076-23384	Bacillomycin synthetase, BmyC protein [*Bacillus amyloliquefaciens* FZB42] CAE11248	
orf8	23374-25200	Putative ABC-transporter TycD [*Brevibacillus brevis*] AAC45931	Extracellular transport of polymyxin?
orf9	25197-26930	Putative ABC-transporter TycE [*Brevibacillus brevis*] AAC45932	
orf10	27221-38977	lichenysin synthetase A [*Bacillus licheniformis* ATCC 14580] AAU22002	Polymyxin biosynthesis?

Table 3.3 The substrate specificity-conferring codes of the adenylation domains derived from ORFs identified in the cosmid inserts of Col-8 and Col-19.

	Adenylation domain	Amino acid residues involved in substrate recognition									Predicated substrates
		235	236	239	278	299	301	322	330	331	
Col-8	*orf6*-A1	D	G	F	L	L	G	L	V	Y	Ile
	orf6-A2	D	V	G	E	I	S	A	I	D	Orn
	orf6-A3	D	V	G	E	I	S	A	I	D	Orn
	orf7-A	D	F	W	N	I	G	M	V	H	Thr
	orf10-A1	D	V	G	E	I	S	S	I	D	Orn
	orf10-A2	D	F	W	N	I	G	M	V	H	Thr
	orf10-A3	D	V	G	E	I	S	S	I	D	Orn
Col-19	*orf7*-A1	D	F	W	N	I	G	M	V	H	Thr
	orf7-A2	D	A	F	W	L	G	C	T	F	Val

[a] The adenylation domains derived from each ORF are numbered according to their order in the deduced amino acid sequence.

Table 3.4 Summary of the proposed functions of ORFs identified within the insert of cosmid Col-19.

ORF	Nucleotide position (start-stop)	Top BLAST hit	Proposed function
orf1	557-2404	Phosphotransferase system, trehalose-specific enzyme II, BC component [*Geobacillus kaustophilus* HTA426] BAD76032	Sugar transport
orf2	2559-3236	Transcriptional regulator GntR family [*Geobacillus thermodenitrificans* NG80-2] ABO67006	Regulation of fusaricidin biosynthesis?
orf3	3400-4196	Two-component response regulator [*Clostridium difficile* 630] CAJ67933	
or4	4387-4737	Acetolactate synthase large subunit [*Saccharopolyspora erythraea* NRRL 2338] CAL99949	Possibly required for acylation of the N-terminal amino acid residue with a fatty acid in the early steps of fusaricidin biosynthesis.
orf5	4944-6002	3-oxoacyl-(acyl-carrier-protein) synthase II [*Carboxydothermus hydrogenoformans* Z-2901] ABB14567	
orf6	5902-6414	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase [*Bacillus cereus* ATCC 14579] AAP09846	
orf7	6508-16257	Bacitracin synthetase 3; BacC [*Bacillus licheniformis*] AAC06348	Fusaricidin biosynthesis

Table 3.4 (continued)

ORF	Nucleotide position (start-stop)	Top BLAST hit	Proposed function
orf10	17810-18916	ABC transporter (substrate-binding protein) [*Geobacillus kaustophilus* HTA426] BAD76195	Export of fusaricidin?
orf11	18949-20499	L-arabinose transport, ATP binding protein [*Geobacillus thermodenitrificans* NG80-2] ABO67160	
orf12	20483-21646	L-arabinose ABC transporter (permease) [*Geobacillus kaustophilus* HTA426] BAD76193	
orf14	23613-24056	Leucine-responsive regulatory protein [*Bacillus thuringiensis serovar israelensis* ATCC 35646] EAO55604	Amino acid metabolism
orf16	27814-28350	Glutamate synthase large subunit [*Halobacillus halophilus*] CAL18237	
orf17	28978-30189	DegT/DnrJ/EryC1/StrS aminotransferase [*Clostridium thermocellum* ATCC 27405] ABN53760	
orf18	31606-32301	dTDP-4-dehydrorhamnose reductase-like [*Synechococcus sp.* CC9902] ABB25060	Cell envelope biogenesis

Table 3.5 Summary of ORFs identified in the fusaricidin biosynthetic gene cluster and the flanking regions.

ORF	Gene	Nucleotide position (start-stop)	Size of the protein product (aa)	Proposed function
1	*gat*	853-1428	191	Glutamine amidotransferase of anthranilate synthase
2	*adl*	1432-2328	298	4-amino-4-deoxychorismate lyase
3	*dps*	2361-3191	276	Dihydropteroate synthase
4	*dna*	3390-3752	120	Dihydroneopterin aldolase
5	*hpk*	3762-4310	182	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase
6	*xre*	4271-4462	63	Transcriptional regulator of Xre family
7	*dus*	4508-5536	342	tRNA-dihydrouridine synthase
8	*greA*	5730-6227	165	Transcription elongation factor
9	*lysRS*	6338-7858	506	Lysyl-tRNA synthetase
10	*ktp*	7949-9286	455	K^+ uptake transporter protein
11	*ymcC*	9579-10112	177	YmcC (unknown function)

Table 3.5 (continued)

ORF	Gene	Nucleotide position (start-stop)	Size of the protein product (aa)	Proposed function
12	*fusG*	10739-11473	244	Enoyl-ACP reductase
13	*fusF*	11534-12946	470	Acyl-CoA ligase
14	*fusE*	12991-14214	407	Aldehyde dehydrogenase
15	*fusD*	14211-15911	566	Acetolactate synthase
16	*fusC*	16118-17176	352	β-ketoacyl synthase
17	*fusB*	17076-17588	170	(3R)-hydroxymyristoyl-ACP dehydratase
18	*fusA*	17682-41408	7908	Peptide synthetase
19	*fusTE*	42549-41926	207	α/β hydrolase
20	*rrnA-16s*	43187-44711		16S ribosomal RNA
21	*rrnA-23s*	45016-47943		23S ribosomal RNA

[a] The proposed function of each ORF is predicted based on the nature of the best matching protein identified by BLAST search.

Table 3.6 The amino acid residues lining the substrate-binding pockets of the adenylation domains in fusaricidin synthetase FusA.

Adenylation domain	Residue positions involved in substrate recognition										Predicted substrate	Corresponding residue in fusaricidin
	235	236	239	278	299	301	322	330	331	517		
FusA-A1	D	F	W	N	I	G	M	V	H	K	Thr	L-Thr
FusA-A2	D	A	F	W	L	G	C	T	F	K	Val	D-Val, D-*allo*-Ile, or D-Ile
FusA-A3	D	A	S	T	L	A	G	V	C	K	3H-Tyr	L-Tyr, L-Phe, L-Val, L-Ile, or L-*allo*-Ile
FusA-A4	D	F	W	N	I	G	M	V	H	K	Thr	D-*allo*-Thr
FusA-A5	D	L	T	K	I	G	E	V	G	K	Asn	D-Asn, or D-Gln
FusA-A6	D	F	P	N	F	C	I	V	Y	K	*	D-Ala

[a] 3H-Tyr represents 3-hydroxy-L-tyrosine.

[b] "*" represents uncertain amino acid.

4. Conclusions, significance and future research

4.1 Conclusions and significance

Fusaricidins are a group of lipopeptide antibiotics produced by *Paenibacillus polymyxa* PKB1, consisting of a guanidinylated β-hydroxy fatty acid linked to a cyclic hexapeptide including four amino acid residues in the D-configuration (Figure 1.3.1A). The antifungal activity of fusaricidins against *Leptosphaeria maculans*, a plant pathogenic fungus causing phoma stem canker (blackleg) disease in canola, makes *P. polymyxa* PKB1 of interest as a potential agent for biocontrol of blackleg disease. To understand the detailed biosynthetic steps involved in fusaricidin production, the complete fusaricidin biosynthetic gene cluster (*fus*) has been cloned and characterized from PKB1 strain. Surprisingly, other than *fusA* encoding the peptide synthetase itself, and genes apparently associated with the synthesis and attachment of the N-terminal guanidino-lipid moiety, no genes for regulation, resistance or transport were found in the cluster. The production of fusaricidins must rely upon these functions being provided by genes located elsewhere in the chromosome.

D-amino acid residues are important elements of nonribosomally produced bioactive peptides. Examination of the fusaricidin gene cluster revealed that two distinct strategies for D-amino acid incorporation are employed in fusaricidin synthetase FusA. The occurrence of D-amino acids in the second, fourth and fifth position of fusaricidins is apparently mediated by epimerization domains present in the corresponding modules, whereas no comparable E domain is found in the sixth module corresponding to D-Ala. Incorporation of a D-amino acid in the C-terminal position of the peptide backbone, as seen in fusaricidin synthetase, is rarely observed in other NRPS systems described so far, but at least one other example exists. δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) synthetase, the first enzyme in the biosynthetic pathways of penicillins and cephalosporins in both eukaryotes and prokaryotes, is a peptide synthetase that incorporates a D-amino acid (D-Val) as the C-terminal residue of peptide products. However, it also contains an additional E domain in the last module, presumably catalyzing racemization of the activated L-Val into its D-isomer. In the study presented in this thesis, it was demonstrated that the sixth A domain of FusA exhibits stereospecificity exclusively for D-Ala, a characteristic reported previously in two eukaryotic NRPS systems, and only demonstrated biochemically using a freestanding A domain of the leinamycin synthetase from a streptomycete (Section 3.6.1.2). Together, both demonstrations of activating a D-amino acid by a recombinant A domain open up the possibility of direct incorporation of D-configured residues into novel peptides as an additional option in combinatorial biosynthesis.

Fusaricidin synthetase is also unusual among bacterial peptide synthetases in that a single enzyme, FusA, is responsible for biosynthesis of the entire fusaricidin peptide, a structural feature more typically associated with fungal peptide synthetases. In bacterial systems, genes encoding NRPSs are generally organized into operons, and biosynthesis of peptide products typically involves two or more peptide synthetase proteins, each consisting of one or more functional modules. In contrast, modules of the fungal NRPSs reside within single polypeptide chains, as in the case of cyclosporin synthetase (Weber *et al.*, 1994). The monocistronic nature of *fusA*, together with the presence of two pathways to incorporate D-amino acids, observed previously only in HC-toxin synthetase from the fungus *Cochliobolus carbonum* (Cheng and Walton, 2000), makes fusaricidin synthetase more similar to fungal than to bacterial peptide synthetases. However, additional sequence analysis, such as G+C content of *fusA* and the flanking regions, is definitely needed before a eukaryotic origin for the fusaricidin synthetase gene could be proposed.

In order to verify the involvement of the single peptide synthetase gene *fusA* in fusaricidin biosynthesis, targeted mutations were created on the chromosome of PKB1. The specific mutagenesis protocol used in *P. polymyxa* PKB1 is based on the PCR targeting system developed by Datsenko and Wanner (2000) for use in *E. coli*, and adapted for mutation of *Streptomyces* (Gust *et al*., 2003). In this approach, chromosomal genes cloned on a cosmid or fosmid vector can be specifically disrupted in *E. coli* by recombination with a PCR-amplified antibiotic resistance cassette flanked by two 39 nt DNA sequences identical to the target gene. The presence of an *oriT* site in the disruption cassette permits conjugation to be used for subsequent transfer of the mutated cosmid/fosmid from *E. coli* to the organism that is the source of the target gene. Compared to the electroporation procedure available for *P. polymyxa* strains, intergeneric transfer of DNA from *E. coli* to *P. polymyxa* by conjugation was found to be much more efficient for PKB1, and therefore the use of conjugation to introduce the gene disruption constructs into *P. polymyxa* was an attractive feature of this approach. In addition, the presence of large homologous DNA pieces flanking the target gene on the cosmid/fosmid clones would greatly increase the recombination frequency.

To adapt the Redirect© technology for *P. polymyxa*, an antibiotic resistance cassette functional in *P. polymyxa* was required. For this purpose, the *cat* gene from a *Staphylococcus* plasmid pC194, one of the very few selectable markers useful in *P.*

polymyxa, was used to construct a special hybrid disruption cassette containing two antibiotic resistance genes, *aac(3)IV* (Apra^R) for selection in *E. coli* and *cat* (Cm^R) for selection in *P. polymyxa*. Although Apra^R can be selected directly in *P. polymyxa*, it was subsequently determined that apramycin in combination with polymyxin B was lethal for *P. polymyxa*. Therefore, inclusion of the *cat* gene in the disruption cassette was necessary to allow use of chloramphenicol for selection of *P. polymyxa* exconjugants in which the chromosomal copy of the target gene was replaced by the disruption cassette via homologous recombination, together with polymyxin B to counterselect *E. coli* donors. The usefulness of this protocol in *P. polymyxa* was demonstrated by the generation of fusaricidin biosynthesis mutants (A4 and A6) in which *fusA* was inactivated by insertion of the gene disruption cassette, and by determination of the *fus* cluster boundaries through gene replacement of two ORFs (*ymcC* and *ktp*) in the flanking region. While the resulting *ymcC* and *ktp* mutants were unaffected in fusaricidin production, the complete loss of antifungal activity in *fusA* mutants A4 and A6 provides evidence that *fusA* is part of the fusaricidin biosynthetic gene cluster and is essential for the production of fusaricidins. To our knowledge, this represents the first reported use of PCR targeting for gene disruption in *Paenibacillus* and provides a valuable new technique for generating specific mutations in this and perhaps other related genera.

4.2 Proposed future research

The study described in this thesis provides evidence of a nonribosomal mechanism for the biosynthesis of fusaricidin in *P. polymyxa* PKB1. Since the *fus* gene cluster does not contain genes needed for regulation of fusaricidin production and product secretion, future work should involve identification of these missing genes, which must reside elsewhere on the chromosome and/or perhaps be shared with another NRPS gene cluster. This could be done by transposon mutagenesis of the PKB1 strain. From mutants that lose the antifungal activity against *L. maculans*, the chromosomal regions adjacent to the transposon could be cloned and sequenced in order to identify new genes that are essential for fusaricidin biosynthesis.

In addition, cloning of the NRPS gene associated with fusaricidin production now makes it possible to undertake genetic manipulation of peptide production, and would potentially increase the antifungal activity of this organism. Fusaricidin is a mixture of at least twelve cyclic depsipeptides, resulting from the relaxed substrate specificity of the individual A domains of fusaricidin synthetase. Bioassay results have shown that the antimicrobial activity of fusaricidin analogs vary depending on the particular amino acids present at three defined positions in the peptide sequence. Previous studies (Beatty and Jensen, 2002) also revealed that the antifungal activity of PKB1 strain against *L. maculans* was mainly attributable to production of a mixture of fusaricidins A and B (Figure 1.3.1B). Therefore, the individual adenylation domains in the second, third and fifth module of FusA should be overexpressed and purified to determine their actual substrate preferences, as has been done for FusA-A6. Recently, alterations of A domain's selectivity have been achieved by point mutations of the specificity-conferring codes within the surfactin and CDA synthetases (Eppelmann *et al.*, 2002, Uguru *et al.*, 2004). This site-directed mutagenesis approach offers the potential to bias the substrate specificities of the relevant fusaricidin synthetase modules, thereby increasing the yield of fusaricidin analogs associated with the greatest antifungal activity. Ultimately, this research would aid in the biocontrol of the fungal pathogen *L. maculans* with enhanced effectiveness and less environmental harm.

Cloning of the polymyxin synthetase gene cluster could be another potential topic of future research. Sequence analysis of cosmid Col-8, obtained from the first SuperCos-1 genomic library of *P. polymyxa* PKB1, revealed the presence of a complete ORF, *orf6*, and two partial ORFs, *orf7* and *orf10*, encoding NRPS modules that presumably synthesize part of the peptide moiety of polymyxin (Section 3.2.3). The involvement of *orf6*, *orf7* or *orf10* in polymyxin production could be tested by creation of targeted mutations in the PKB1 chromosome using the PCR targeting mutagenesis protocol developed in this study, and then a DNA fragment corresponding to one of these three ORFs could be used as a probe to screen the pSMART-FOS genomic library in an effort to isolate fosmid clones that may contain peptide synthetase genes from the polymyxin biosynthetic gene cluster.[1][2]

174

---

[1] *A version of this chapter has been published. Li et al. 2007. Applied and Environmental Microbiology. 73: 3480-3489.*

[2] *A version of this chapter has been submitted for pulication. Li and Jensen 2007.*

5. References

Ash, C., Priest, F.G., and Collins, M.D. (1993) Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. *Antonie Leeuwenhoek* **64**: 253-260

Balibar, C.J., Vaillancourt, F.H., and Walsh, C.T. (2005) Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol* **12**: 1189-1200.

Beatty, P.H., and Jensen, S.E. (2002) *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. *Can J Microbiol* **48**: 159-169.

Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**: 486-489.

Bergendahl, V., Linne, U., and Marahiel, M.A. (2002) Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur J Biochem* **269**: 620-629.

Bezzate, S., Aymerich, S., Chambert, R., Czarnes, S., Berge, O., and Heulin, T. (2000) Disruption of the *Paenibacillus polymyxa* levansucrase gene impairs its ability to aggregate soil in the wheat rhizosphere. *Environ Microbiol* **2**: 333-342.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., and Stubbs, M.T. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**: 301-310.

Burmester, J., Haese, A., and Zocher, R. (1995) Highly conserved N-methyltransferases as an integral part of peptide synthetases. *Biochem Mol Biol Int* **37**: 201-207.

Cane, D.E., and Walsh, C.T. (1999) The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem Biol* **6**: R319-325.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**: 211-224.

Challis, G.L., and Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* **14**: 748-756.

Cheng, Y.Q., and Walton, J.D. (2000) A eukaryotic alanine racemase gene involved in cyclic peptide biosynthesis. *J Biol Chem* **275**: 4906-4911.

Clugston, S.L., Sieber, S.A., Marahiel, M.A., and Walsh, C.T. (2003) Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. *Biochemistry* **42**: 12095-12104.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J* **16**: 4174-4183.

Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G., and van Sinderen, D. (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol Microbiol* **8**: 821-831.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Pro Natl Acad Sci USA* **97**: 6640-6645.

de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C., and Grandi, G. (1997) Engineering of peptide synthetases. Key role of the thioesterase-like domain for efficient production of recombinant peptides. *J Biol Chem* **272**: 25304-25309.

Dieckmann, R., Lee, Y.O., van Liempt, H., von Dohren, H., and Kleinkauf, H. (1995) Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS Lett* **357**: 212-216.

Dijksterhuis, J., Sanders, M., Gorris, L.G., and Smid, E.J. (1999) Antibiosis plays a role in the context of direct interaction during antagonism of *Paenibacillus polymyxa* towards *Fusarium oxysporum*. *J Appl Microbiol* **86**: 13-21.

Dittmann, J., Wenger, R.M., Kleinkauf, H., and Lawen, A. (1994) Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor. *J Biol Chem* **269**: 2841-2846.

Doekel, S., and Marahiel, M.A. (2000) Dipeptide formation on engineered hybrid peptide synthetases. *Chem Biol* **7**: 373-384.

Du, L., Chen, M., Sanchez, C., and Shen, B. (2000) An oxidation domain in the *Blm*III non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. *FEMS Microbiol Lett* **189**: 171-175.

Du, L., Sanchez, C., and Shen, B. (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab Eng* **3**: 78-95.

Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F., and Vater, J. (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Pro Natl Acad Sci USA* **96**: 13294-13299.

Ehmann, D.E., Trauger, J.W., Stachelhaus, T., and Walsh, C.T. (2000) Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol* **7**: 765-772.

Elsner, A., Engert, H., Saenger, W., Hamoen, L., Venema, G., and Bernhard, F. (1997) Substrate specificity of hybrid modules from peptide synthetases. *J Biol Chem* **272**: 4814-4819.

Eppelmann, K., Stachelhaus, T., and Marahiel, M.A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**: 9718-9726.

Eriani, G., Dirheimer, G., and Gangloff, J. (1990) Aspartyl-tRNA synthetase from *Escherichia coli*: cloning and characterisation of the gene, homologies of its translated amino acid sequence with asparaginyl- and lysyl-tRNA synthetases. *Nucleic Acids Res* **18**: 7109-7118.

Ferrari, F.A., Nguyen, A., Lang, D., and Hoch, J.A. (1983) Construction and properties of an integrable plasmid for *Bacillus subtilis*. *J Bacteriol* **154**: 1513-1515.

Gaisser, S., and Hughes, C. (1997) A locus coding for putative non-ribosomal peptide/polyketide synthase functions is mutated in a swarming-defective Proteus mirabilis strain. *Mol Gen Genet* **253**: 415-427.

Gaitatzis, N., Hans, A., Muller, R., and Beyer, S. (2001) The *mtaA* gene of the myxothiazol biosynthetic gene cluster from *Stigmatella aurantiaca* DW4/3-1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases. *J Biochem* **129**: 119-124.

Gehring, A.M., Mori, I., Perry, R.D., and Walsh, C.T. (1998) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**: 11637-11650.

Gocht, M., and Marahiel, M.A. (1994) Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. *J Bacteriol* **176**: 2654-2662.

Grunewald, J., Sieber, S.A., and Marahiel, M.A. (2004) Chemo- and regioselective peptide cyclization triggered by the N-terminal fatty acid chain length: the recombinant cyclase of the calcium-dependent antibiotic from *Streptomyces coelicolor*. *Biochemistry* **43**: 2915-2925.

Guenzi, E., Galli, G., Grgurina, I., Gross, D.C., and Grandi, G. (1998) Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J Biol Chem* **273**: 32857-32863.

Gugel, R.K., and Petrie, G.A. (1992) History, occurrence, impact, and control of blackleg of rapeseed. *Can J Plant Pathol* **14**: 36-45.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* **100**: 1541-1546.

Haese, A., Schubert, M., Herrmann, M., and Zocher, R. (1993) Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. *Mol Microbiol* **7**: 905-914.

Haese, A., Pieper, R., von Ostrowski, T., and Zocher, R. (1994) Bacterial expression of catalytically active fragments of the multifunctional enzyme enniatin synthetase. *J Mol Biol* **243**: 116-122.

Hall, R. (1992) Epidemiology of blackleg of oilseed rape. *Can J Plant Pathol* **14**: 46-55.

Han, Y.W. (1989) Levan production by *Bacillus polymyxa*. *J Ind Microbiol* **4**: 447-451.

Heulin, T., Berge, O., Mavingui, P., Gouzou, L., Hebbar, K.P., and Balandreau, J. (1994) *Bacillus polymyxa* and *Rahnella aquatilis*, the dominant N_2-fixing bacteria associated with wheat rhizosphere in French soils. *Eur J Soil Biol* **30**: 35-42.

Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H., and Zocher, R. (1994) Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *J Biol Chem* **269**: 12710-12714.

Holl, F.B., Chanway, C.P., Turkington, R., and Radley, R.A. (1988) Response of crested wheatgrass (*Agropyron cristatum* L), perennial ryegrass (*Lolium perenne*) and white clover (*Trifolium repens* L) to inoculation with *Bacillus polymyxa*. *Soil Biol Biochem* **20**: 19-24.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, K., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M., and Schrempf, H. (1985) Genetic manipulation of *Streptomyces*: a laboratory manual. John Innes Foundation, Norwich, UK.

Horinouchi, S., and Weisblum, B. (1982) Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J Bacteriol* **150**: 815-825.

Itaya, M., Kondo, K., and Tanaka, T. (1989) A neomycin resistance gene cassette selectable in a single copy state in the *Bacillus subtilis* chromosome. *Nucleic Acids Res* **17**: 4410.

Ito, M., and Koyama, Y. (1972a) Jolipeptin, a new peptide antibiotic. I. Isolation, physico-chemical and biological characteristics. *J Antibiot* **25**: 304-308.

Ito, M., and Koyama, Y. (1972b) Jolipeptin, a new peptide antibiotic. II. The mode of action of jolipeptin. *J Antibiot* **25**: 309-314.

Janssen, G.R., and Bibb, M.J. (1993) Derivatives of pUC18 that have *Bgl*II sites flanking a modified multiple cloning site and that retain the ability to identify recombinant clones by visual screening of *Escherichia coli* colonies. *Gene* **124**: 133-134.

Jisha, M.S., and Alagawadi, A.R. (1996) Nutrient uptake and yield of sorghum (*Sorghum bicolor* L Moench) inoculated with phosphate solubilizing bacteria and cellulolytic fungus in a cotton stalk amended vertisol. *Microbiol Res* **151**: 213-217.

Jurtshuk, R.J., Blick, M., Bresser, J., Fox, G.E., and Jurtshuk, P., Jr. (1992) Rapid in situ hybridization technique using 16S rRNA segments for detecting and differentiating the closely related gram-positive organisms *Bacillus polymyxa* and *Bacillus macerans*. *Appl Environ Microbiol* **58**: 2571-2578.

Kajimura, Y., and Kaneda, M. (1996) Fusaricidin A, a new depsipeptide antibiotic produced by *Bacillus polymyxa* KT-8. Taxonomy, fermentation, isolation, structure elucidation and biological activity. *J Antibiot* **49**: 129-135.

Kajimura, Y., and Kaneda, M. (1997) Fusaricidins B, C and D, new depsipeptide antibiotics produced by *Bacillus polymyxa* KT-8: isolation, structure elucidation and biological activity. *J Antibiot* **50**: 220-228.

Katz, E., and Demain, A.L. (1977) Peptide antibiotics of *Bacillus* - chemistry, biogenesis, and possible functions. *Bacteriol Rev* **41**: 449-474.

Keating, T.A., Marshall, C.G., and Walsh, C.T. (2000) Reconstitution and characterization of the *Vibrio cholerae* vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **39**: 15522-15530.

Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* **9**: 522-526.

Kessler, N., Schuhmann, H., Morneweg, S., Linne, U., and Marahiel, M.A. (2004) The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J Biol Chem* **279**: 7413-7419.

Khangura, R.K., and Barbetti, M.J. (2001) Prevalence of blackleg (*Leptosphaeria maculans*) on canola (*Brassica napus*) in Western Australia. *Aust J Exp Agric* **41**: 71-80.

Kharbanda, P.D. (1992) Performance of fungicides to control blackleg of canola. *Can J Plant Pathol* **14**: 169-176.

Kharbanda, P.D., and Dahiya, J.S. (1990) A metabolite of *Penicillium verrucosum* inhibitory to growth of *Leptosphaeria maculans* and *Rhizoctonia solani*. *Can J Plant Pathol* **12**: 335 (Abstract).

Kieleczawa, J. (ed) (2005) DNA sequencing: optimizing the process and analysis. Jones and Bartlett, Sudbury, MA.

Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A., and Walsh, C.T. (2001) Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* **40**: 7099-7108.

Kohli, R.M., Walsh, C.T., and Burkart, M.D. (2002) Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**: 658-661.

Konz, D., Doekel, S., and Marahiel, M.A. (1999) Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J Bacteriol* **181**: 133-140.

Konz, D., and Marahiel, M.A. (1999) How do peptide synthetases generate structural diversity? *Chem Biol* **6**: R39-48.

Kuroda, J., Fukai, T., Konishi, M., Uno, J., Kurusu, K., and Nomura, T. (2000) LI-F antibiotics, a family of antifungal cyclic depsipeptides produced by *Bacillus polymyxa* L-1129. *Heterocycles* **53**: 1533-1549.

Kurusu, K., Ohba, K., Arai, T., and Fukushima, K. (1987) New peptide antibiotics LI-F03, F04, F05, F07, and F08, produced by *Bacillus polymyxa* .1. Isolation and Characterization. *J Antibiot* **40**: 1506-1514.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **3**: 923-936.

Li, J., Beatty, P.K., Shah, S., and Jensen, S.E. (2007) Use of PCR-targeted mutagenesis to disrupt production of fusaricidin-type antifungal antibiotics in *Paenibacillus polymyxa*. *Appl Environ Microbiol* **73**: 3480-3489.

Li, J., and Jensen, S.E. (2007) Nonribosomal biosynthesis of fusaricidins by *Paenibacillus polymyxa* PKB1 involves direct incorporation of a D-amino acid. (submitted).

Linne, U., and Marahiel, M.A. (2000) Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39**: 10439-10447.

Linne, U., Doekel, S., and Marahiel, M.A. (2001) Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. *Biochemistry* **40**: 15824-15834.

Lipmann, F. (1971) Attempts to map a process evolution of peptide biosynthesis. *Science* **173**: 875-884.

Lipmann, F., Gevers, W., Kleinkauf, H., and Roskoski, R., Jr. (1971) Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. *Adv Enzymol Relat Areas Mol Biol* **35**: 1-34.

Lipmann, F. (1973) Non-ribosomal polypeptide synthesis. *Acta Endocrinol Suppl (Copenh)* **180**: 294-300.

Lipmann, F. (1980) Bacterial production of antibiotic polypeptides by thiol-linked synthesis on protein templates. *Adv Microb Physiol* **21**: 227-266.

Luo, L., Kohli, R.M., Onishi, M., Linne, U., Marahiel, M.A., and Walsh, C.T. (2002) Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **41**: 9184-9196.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**: 61-68.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* **97**: 2651-2674.

Marck, C. (1988) 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res* **16**: 1829-1836.

Mavingui, P., and Heulin, T. (1994) In-vitro chitinase and antifungal activity of a soil, rhizosphere and rhizoplane population of *Bacillus polymyxa*. *Soil Biology & Biochemistry* **26**: 801-803.

May, J.J., Kessler, N., Marahiel, M.A., and Stubbs, M.T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci USA* **99**: 12120-12125.

Mcgee, D.C., and Petrie, G.A. (1978) Variability of *Leptosphaeria maculans* in relation to blackleg of oilseed rape. *Phytopathology* **68**: 625-630.

Miao, V., Coeffet-Legal, M.F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C.J., Wrigley, S.K., and Baltz, R.H. (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**: 1507-1523.

Miller, D.A., and Walsh, C.T. (2001) Yersiniabactin synthetase: probing the recognition of carrier protein domains by the catalytic heterocyclization domains, Cy1 and Cy2, in the chain-initiating HWMP2 subunit. *Biochemistry* **40**: 5313-5321.

Mootz, H.D., and Marahiel, M.A. (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* **179**: 6843-6850.

Mootz, H.D., and Marahiel, M.A. (1999) Design and application of multimodular peptide synthetases. *Curr Opin Biotechnol* **10**: 341-348.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2000) Construction of hybrid peptide synthetases by module and domain fusions. *Proc Natl Acad Sci U S A* **97**: 5848-5853.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **3**: 490-504.

Moyne, A.L., Cleveland, T.E., and Tuzun, S. (2004) Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. *FEMS Microbiol Lett* **234**: 43-49.

Nakajima, N., Koyama, Y., and Chihara, S. (1972) New antibiotic, gatavalin .1. Isolation and Characterization. *J Antibiot* **25**: 243-246.

O'Connor, T.J., Kanellis, P., and Nodwell, J.R. (2002) The *ramC* gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Mol Microbiol* **45**: 45-57.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**: 2411-2423.

Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**: 204-211.

Pavela-Vrancic, M., Pfeifer, E., Schroder, W., von Dohren, H., and Kleinkauf, H. (1994) Identification of the ATP binding site in tyrocidine synthetase 1 by selective modification with fluorescein 5'-isothiocyanate. *J Biol Chem* **269**: 14962-14966.

Petersen, D.J., Srinivasan, M., and Chanway, C.P. (1996) *Bacillus polymyxa* stimulates increased *Rhizobium etli* populations and nodulation when co-resident in the rhizosphere of *Phaseolus vulgaris*. *FEMS Microbiol Lett* **142**: 271-276.

Pichard, B., Larue, J.P., and Thouvenot, D. (1995) Gavaserin and saltavalin, new peptide antibiotics produced by *Bacillus polymyxa*. *FEMS Microbiol Lett* **133**: 215-218.

Quigley, N.B., Mo, Y.Y., and Gross, D.C. (1993) SyrD is required for syringomycin production by *Pseudomonas syringae* pathovar syringae and is related to a family of ATP-binding secretion proteins. *Mol Microbiol* **9**: 787-801.

Rapaport, E., Yogeeswaran, G., Zamecnik, P.C., and Remy, P. (1985) Covalent modification of phenylalanyl-tRNA synthetase with phenylalanine during the amino acid activation reaction catalyzed by the enzyme. *J Biol Chem* **260**: 9509-9512.

Rimmer, S.R., and Vandenberg, C.G.J. (1992) Resistance of oilseed *Brassica* spp to blackleg caused by *Leptosphaeria maculans*. *Can J Plant Pathol* **14**: 56-66.

Roongsawang, N., Hase, K., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2003) Cloning and characterization of the gene cluster encoding arthrofactin synthetase from *Pseudomonas* sp. MIS38. *Chem Biol* **10**: 869-880.

Rosado, A., and Seldin, L. (1993) Isolation and partial characterization of a new linear DNA plasmid isolated from *Bacillus polymyxa* SCE2. *J Gen Microbiol* **139**: 1277-1282.

Ruttenberg, M.A., and Mach, B. (1966) Studies on amino acid substitution in the biosynthesis of the antibiotic polypeptide tyrocidine. *Biochemistry* **5**: 2864-2869.

Salowe, S.P., Marsh, E.N., and Townsend, C.A. (1990) Purification and characterization of clavaminate synthase from *Streptomyces clavuligerus*: an unusual oxidative enzyme in natural product biosynthesis. *Biochemistry* **29**: 6499-6508.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M.A., Kruft, V., and Wittmann-Liebold, B. (1991) An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J Biol Chem* **266**: 23135-23141.

Schneider, A., and Marahiel, M.A. (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch Microbiol* **169**: 404-410.

Schneider, T.L., Shen, B., and Walsh, C.T. (2003) Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**: 9722-9730.

Scholz-Schroeder, B.K., Soule, J.D., and Gross, D.C. (2003) The sypA, sypS, and sypC synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. syringae B301D. *Mol Plant Microbe Interact* **16**: 271-280.

Schwarzer, D., Finking, R., and Marahiel, M.A. (2003) Nonribosomal peptides: from genes to products. *Nat Prod Rep* **20**: 275-287.

Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci USA* **99**: 14083-14088.

Scott-Craig, J.S., Panaccione, D.G., Pocard, J.A., and Walton, J.D. (1992) The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. *J Biol Chem* **267**: 26044-26049.

Shida, O., Takagi, H., Kadowaki, K., Nakamura, L.K., and Komagata, K. (1997a) Emended description of *Paenibacillus amylolyticus* and description of *Paenibacillus illinoisensis* sp. nov. and *Paenibacillus chibensis* sp. nov. *Int J Syst Bacteriol* **47**: 299-306.

Shida, O., Takagi, H., Kadowaki, K., Nakamura, L.K., and Komagata, K. (1997b) Transfer of *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*. *Int J Syst Bacteriol* **47**: 289-298.

Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S., and Muller, R. (1999) New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J Biol Chem* **274**: 37391-37399.

Silakowski, B., Kunze, B., Nordsiek, G., Blocker, H., Hofle, G., and Muller, R. (2000) The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur J Biochem* **267**: 6476-6485.

Sogn, J.A. (1976) Structure of peptide antibiotic polypeptin. *J Med Chem* **19**: 1228-1231.

Stachelhaus, T., Schneider, A., and Marahiel, M.A. (1995) Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**: 69-72.

Stachelhaus, T., Huser, A., and Marahiel, M.A. (1996) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem Biol* **3**: 913-921.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J Biol Chem* **273**: 22773-22781.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**: 493-505.

Stachelhaus, T., and Walsh, C.T. (2000) Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **39**: 5775-5787.

Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., Jacques, P., Thonart, P., and Vater, J. (1999) Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chem Biol* **6**: 31-41.

Tang, G.L., Cheng, Y.Q., and Shen, B. (2004) Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem Biol* **11**: 33-45.

Tang, G.L., Cheng, Y.Q., and Shen, B. (2007) Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from *Streptomyces atroolivaceus* S-140. *J Biol Chem* **282**: 20273-20282.

Tosato, V., Albertini, A.M., Zotti, M., Sonda, S., and Bruschi, C.V. (1997) Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. *Microbiology* **143**: 3443-3450.

Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A., and Walsh, C.T. (2000) Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**: 215-218.

Trauger, J.W., Kohli, R.M., and Walsh, C.T. (2001) Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry* **40**: 7092-7098.

Tseng, C.C., Bruner, S.D., Kohli, R.M., Marahiel, M.A., Walsh, C.T., and Sieber, S.A. (2002) Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* **41**: 13350-13359.

Tsuge, K., Akiyama, T., and Shoda, M. (2001) Cloning, sequencing, and characterization of the iturin A operon. *J Bacteriol* **183**: 6265-6273.

Turgay, K., Krause, M., and Marahiel, M.A. (1992) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6**: 529-546.

Turgay, K., and Marahiel, M.A. (1994) A general approach for identifying and cloning peptide synthetase genes. *Pept Res* **7**: 238-241.

Uguru, G.C., Milne, C., Borg, M., Flett, F., Smith, C.P., and Micklefield, J. (2004) Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. *J Am Chem Soc* **126**: 5032-5033.

Valser-Genser, M., and Eskin, N.A. (1982) Canola oil, properties and performance. Canola Coucil of Canada. Manitoba.

Vieira, J., and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods Enzymol* **153**: 3-11.

von Dohren, H., Keller, U., Vater, J., and Zocher, R. (1997) Multifunctional peptide synthetases. *Chem Rev* **97**: 2675-2706.

Walker, S., Chen, L., Hu, Y., Rew, Y., Shin, D., and Boger, D.L. (2005) Chemistry and biology of ramoplanin: a lipoglycodepsipeptide with potent antibiotic activity. *Chem Rev* **105**: 449-476.

Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E., and Flugel, R.S. (1997) Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr Opin Chem Biol* **1**: 309-315.

Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* **5**: 525-534.

Walton, J.D. (1987) Two enzymes involved in biosynthesis of the host-selective phytotoxin HC-toxin. *Proc Natl Acad Sci U S A* **84**: 8444-8447.

Wang, Y.P., Nowak, G., Culley, D., Hadwiger, L.A., and Fristensky, B. (1999) Constitutive expression of pea defense gene DRR206 confers resistance to blackleg (*Leptosphaeria maculans*) disease in transgenic canola (*Brassica napus*). *Molr Plant-Microbe Interact* **12**: 410-418.

Weber, G., Schorgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* **26**: 120-125.

Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**: 407-418.

Williams, R.H., and Fitt, B.D.L. (1999) Differentiating A and B groups of *Leptosphaeria maculans*, causal agent of stem canker (blackleg) of oilseed rape. *Plant Pathol* **48**: 161-175.

Yeh, E., Kohli, R.M., Bruner, S.D., and Walsh, C.T. (2004) Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem* **5**: 1290-1293.

Yin, X., and Zabriskie, T.M. (2006) The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* **152**: 2969-2983.

Zocher, R., Nihira, T., Paul, E., Madry, N., Peeters, H., Kleinkauf, H., and Keller, U. (1986) Biosynthesis of cyclosporin A: partial purification and properties of a multifunctional enzyme from *Tolypocladium inflatum*. *Biochemistry* **25**: 550-553.

6. Appendices

Appendix A: The nucleotide sequence of the insert DNA of cosmid Col-8

Sequence 39063 bp; 8943 A; 10051 C; 11619 G; 8450 T

```
   1 gatcctgttc ttcaagcaag accgccaatt cctcgatctt cttcagccag cttcgcgacg
  61 gagtggcgtt gacagcaagt acgttccaag tgggtaaatt cccattcatg atgttcggtt
 121 tgccggcagc agcgcttgca atgtaccatg aagcaagacc agaacacaaa aaatatgtag
 181 caggtatcat gggttcagca gcattgactt cgttcttgac aggtattacg gagccactgg
 241 aatttccctt cctgttcgta gctccactgc tgtttgcagt acactgtata ttcgcaggtc
 301 tgtcatttat gaccatgcaa attttacatg tcaaaatcgg tatgaccttc tcaggcggat
 361 ttattgactt cctgatttcc gggattattc caacccgtac gccttggtgg tatgtaatca
 421 tcgtcggtct gattttggca gtgatttact acttcggatt ccggttcatc atacggaagt
 481 tcaatttgaa aacaccagga cgtgaagaag ctacagaaaa tgtggatgac gcagcagccg
 541 gtagtagtga tgagctcccg cacaacattt tggcagcttt cggtggttca tcgaatatca
 601 agcatctgga tgcttgtatc acacgtcttc ggatcgaggt taatgagaaa tcgaacgttg
 661 acaaagctcg tctgaaacaa cttggcgcat cgggtgtcct tgaagtcggc aacaacgtac
 721 aggcaatttt cggtacacgt tccgacacca ttaaatctca aatggctgat atcatggctg
 781 gacgtacacc tgcaccagct ccggcgcaac ctgctccaca agaggagaag gaagctggag
 841 aagaacaaca aaccattatt gtagaagata ttgtaatgcc tgttaacggc gaattggttg
 901 atatctcgac ggttcccgat ccccgcgtaa atatcttctg tgttctcacg gaaaataacc
 961 atgtctacca attcaggacg tttcactgga gaaggaacac cgttaaaata acgaacagga
1021 cgcagacata cgtataaatc cagttcttgc cgcaatgcca cattcaacga acggatacca
1081 ccgccaatag gtgtcgtaag tggacctttg atggctacaa tatactcacg gatggcttcc
1141 aaagtatcgt tagggagcca ttcgccgtat gtattgaaag ctttctcacc ggcgaatact
1201 tcataccagg caatcttttt gctgccatta tatgcttttt ctacagctgc gtccagtacg
1261 cgcttggaag ctttccagat atcgcgaccc gtaccgtcac cttcgataaa cggaattacc
1321 ggatggttcg gaaccggtag ttgaccgtta tcaatcgtaa ttttttcgcc ttcagttggg
1381 agctcaaact tttcaaattt cgccataagt tccttttcct cctcagatat gaatgggtac
1441 aggacgaaag acatgatctg actcttcggt aaatccctga tatgacccat ggaagccttg
1501 acgatatatt tactgcctaa gtatttgccg atcgaagaac aagttccgct gcatttgctc
1561 cgtcgtaaac aacccggcat acaccacagg ctgcacagtt tccaagtcat tcacgggaat
1621 gtagtactga ttatccacca gctcgtacaa cggggcattt tcaccgaaat tcacatgctg
1681 ccgaatgtcc tctaccgtca tgtcgaactg gttcgcctca tacaccacat cgcctctggc
1741 actgaaaaag agcgcgtgtg gcttggcatt tttttcgtct atatagatcc gcagaccgta
1801 cagttcattc aggcggttcg agctgggttc aggccgaata cggtcgttat tttgaaggat
1861 gaaggacagt cggagattgc agagccggtt tcttttatcc gtgattacga tctcaaagtc
1921 agtcttgctc acaacattca tgacaagttt ggcggaagta ttgccctgtt caatgagtac
1981 ggacctacgg aaaccgttgt aggctgcatg atccatatgt atgatgcaga ccgggatagc
2041 agaagatccg ttcccatcgg cgtgccttcc gataacgtgg ctgtatattt gctggatgcg
2101 catttaaagc ctgttcctat gcaggtaaaa ggagaaatct acatttccgg tcatggagtg
```

```
2161 gcaagaggat atctgaacag accggaactc acaaaagaaa ggtttctcga taatcctttt
2221 gttcctggcg aacgaatgta taaaaccggt gacatgggga tacggctgga aaatggcctg
2281 atcgaatacc ttggccgaaa tgatcaccag gttaaaatca gaggctttcg tatcgagctg
2341 ggtgaaattg aaggagcttt atcctcctat cctgacatcc agcatgctgt cgtaaacgtt
2401 gtggaaacag acgatgcgaa tagatatccc gttttgtgtg cctattatgt atctgatcgt
2461 gctatctctc ccaagcttct caaagattat ttacatgctt gtcttcccca ttatatggta
2521 ccaagccaca tactccagct ggaaatcatc ccgcttacat ccaatggcaa ggtggacaag
2581 caaaagcttc ctgtaccaga cgtgtcctct ggattaactg acatatatga ggaagcccat
2641 aatacggtgg aaagcattct cgtgcaggtt tgggaagaga tgttccatac gtctatcata
2701 ggtatccacg ataatttctt tgcactgggc ggagactcta ttaaagcgat ccagatgaca
2761 tccaaattaa ataactatga tttggaagtg ggtgtacagg atgttttagc gcatcccagt
2821 atagcagagc ttgctcatta cacacgccaa aagagccgac tgtacgattc ttcagctccc
2881 gtcagcggag aaattcagcc tacaccgatc tccttgtggt ttatggagca gtcatttcat
2941 aatcctcatc actataatca atcggtcttg cttctgctga aaaatgaaat ggatcgttct
3001 gctctagaaa aggcatttca caaagtaatt gaacatcacg atgctttaag gatgagcata
3061 cacccggatg gcaagttatt ttataacccg gaactcactg caatcccatt caagctggat
3121 tgttatgaca tgactcctat ttcagcagag gctcggcagc atgaattcaa taagctggtg
3181 catagcttgc aatccggttt cgatctgtcc caccgtcttc ctattcgagc agctatattc
3241 gatcacggac aagaaacctg ggaattgttc atcactgcgc atcatctggt tgtagatgga
3301 gtttcctgga gaattatatt ggaggatgta ttggaggctt atcatgcctt agaagaaggg
3361 aaggaggtac attttcaag aaaaacagcc tctgttcaaa cctttgccaa ggagcttcat
3421 cattattcca ctactgcgga tctacgggag gagcttgatt tttggaatga aatggagggc
3481 ggcaacgacc attttccagc attcatcaca agccagccat ccactgccgt aaactatgct
3541 tcatgtacag tcgtcgaagg agaactttcc cctgcccata cgcacaaact gctcaccaag
3601 gcgaatcagg tctatcatac cgaacctgtg gatttactgc tggctgccct ggcgttaagt
3661 gtaaaggaat ggattggatt ggacgatttc acttatgagg ttgagcatca cggaaggagc
3721 ctggatcacg tagacgtatc cagaacggtt ggctggttta cggcgttaca tcccttacgt
3781 attcaagtgc caggtacaga aataggcagt gtcattgcat atgtcaagga actgcgccga
3841 cgcattcccc aacagggtat cggttatggc attttaaaat acatgctagg agctattgaa
3901 aaaagtaaaa agattcgtcc cctacgcttc aattatctgg gacagttcga tcacgaaagt
3961 cagagttcag actatgtata taagcatact ccaggccttc aagatgtgga ctacgccaat
4021 catctaacgg ctggcatcga aatcaacagc ctgattgtta acgaccagtt ggtagtccat
4081 atgatgtaca gcacaacaat gtttgaaaaa gaaagtataa ccgaatttat gcatatgtat
4141 gtccaaactc tacaccgtgt gattgatttt acacttagtc agcagaatgt tcacttcact
4201 ccttcggact ttgaatctgc tcggattatc caagaggata ttgaattgct tttcgaataa
4261 gagtgcaata agggcaggcg gttgacttca accgcctgtc taattagact tcaggcgtaa
4321 ctactgtact tttttcatat aaaggtggaa taggtagtga atgagttctt tcgaagatca
4381 tttggttgtt ttaaacacac ccagacggtt cagtacgacc agtaagcgat aaaaatcgta
4441 actgcttggc agtttcgcat ccacaatcgg gcttgccgga ttaaatgcaa cagcggcagc
4501 catcgcttcc ttcgcccagg tcggtacagg catagcctga cgcgcctcca gttcagctac
4561 ccgctgatcc accttttttaa tataagcgct ctgttcattc aacgtctgct tgagcacatc
```

```
4621 tttgctgttt tcaagggcgg cgacctgctg ctgcagggcc ttaaataccg ctttttcctc
4681 tgctgtcatt ggctcatccc cctttctcc tacaaatgct tggcgtagct cggtctctgt
4741 tccgacgtat acattcaggt caaccggccc ttgaattcca tcaacccgtc cactatcgct
4801 gtattgccaa aagttccagc gcttccatgc cgtggtgtcg ccgggcaccc gtgtactata
4861 acgcgcaatc cacagcttat aactacccaa tggggctttg aaatgggcag caaaagcatt
4921 ccctgtatat accatcggct tacggccgct tacccgttca aattccgtta caaatgccaa
4981 cgcaacctca tgaatcgcgg ccgagctaag ccctcccgaa ttattttcgt aatccatcac
5041 cgctggcaag tccagcgctt tcgcaccccc cacctgatcc agcacctcgg cgaaatgcct
5101 ggcttccgct tttgctgcgt ttacactagt tgcatctaca aaatggtaag ctcccaggag
5161 aatcccggct gccttgatcc cttttgtatt cgtgataaat gtcggatcga cgtaacgctg
5221 tccttgactc gctttaataa aggcaaacga tttaccatcg gctcgtacac gtttccaatc
5281 aatcgttccc tgatatcggg atacgtcaat gccttcggca gcgttagggc tgcgattctg
5341 cataaggcgt tccctcctta ttattcataa tgttgtagct cactctgtct ctgttatcca
5401 tatgcttgga ccatccatac ggtctgggct tttccacttt tataccgaaa aaagggcaaa
5461 aaggtagtct gcctattttg gtagtggcca ctagtggccc tttatgaagc tttattttct
5521 ttgtcggaag ggtcggttgg ctttattttt tctccactcg acttagactc aaaaatagct
5581 accgcattcc gcaaaatatc cggcatcggc actcccattc tgcccacatt ttcaataatg
5641 gacaacagct cgttcgctaa ataaaagaaa acgacggcat cccggaagta atgtgcatcc
5701 ccgagaatac cgtcgattag atgagctacc gtaatgagca aaaacacggt aaccttgcga
5761 aaaataccgt aatacccctg gcgacttttt agctccccat taatccaggc cgctgcccac
5821 ccggtgaaaa agtcgatcac gaccatccac cacagcaacg tcaacattgt ggtccacccc
5881 ccaaaaaaat aaccgatcca atctatgaaa aagacgccgg gatgactccg aaaggcgtct
5941 tttttggtt tggcagcagc taataaagga catgatgaac atacatagca ttctcgtgca
6001 taaactctta aaaagcgaat gacaaaagga gagagcgagc gtatgatgtc ctttatttta
6061 gggtgtgtga ttgtctattt actcgtagct gtcagtagcg aaaaagcaga ttcaattgtg
6121 gagttaaatg attctgtgaa ataaattttt ttattgaaaa tcatgcattt tctgatagga
6181 tagagttaaa tattggaatc ccaaaggaga atgcatgctg gaattttcat ttgaaatcat
6241 cgacgaatcg aaaatcaaca ttgtatatca gtatggaagc agtagcttca attttaattt
6301 atttttaac tacggtgtgt ggacgctgca tcctttgac gggattctgc tgcagaacaa
6361 ggaaatgtgc cgtctgatct gccaaatgtt gaatatcgat gtagcctagt cggtgaaaga
6421 gctcctcctc gcgcagatgg ttcggatcgg cctttttccg ggttcgtggc ggctcaatga
6481 ccagcacaac tccaaaacgt ttggaaggta cgttaaaccg ttgctgccgc aagtagagga
6541 aaaacgggtc aaggtcaagc ggtataggct ccccctcgtc ctgttcgtcc gctcgttcct
6601 tgtcatagtg gaataccagc cccacatctt cgatgatcgc cactttgccg ccaacttatg
6661 cggcttattt gaattcagac cgcctgccga gtctgagcag gctcatcacg gcgggctccg
6721 ccgtttcagc cgaattcgtg cagcagtgga aagacaaggt acagtactat aacgcgtacg
6781 gtccgacgga agcctcgatc gccacatccg tgtgggcggc gtctacgtat gacaccgaac
6841 gcagagcgat accgatcgga cgtccgatct ggaaccatcg cttgtatata ctgggcgcac
6901 agaatcagct ggcgcctatc ggagtggagg gggaattgtg catcgcgggc gtcggcctcg
6961 cccggggata ttggaaccgg ccggatctga cggcggagaa attcatcgaa cacccgtttg
7021 agcccggaga gagaatgtac cggacgggcg acttggcccg ctggctgccg gacggaaata
```

```
7081 ttgagtactt gggccggatt gaccatcagg tgaaaatccg tggattccgg atcgagatcg
7141 gcgagattga agagcagctt ctgaagatcg actccgtgca ggagacgatc gtaatcgcgc
7201 gggaaggcaa aagcgggcaa gaactgtgcg cttatctggt cgcgggccgc ccgcttacgc
7261 tcggcgagct gagaagcgcg ctggcgcaaa aattgccgaa ttacatgatt ccggcgcatt
7321 ttgttcagct tccgcggatg ccgctcacgc cgaacgacaa aatcgaccgc aaggctttgc
7381 ccgctccgga aggaaacgcg ctgaccggcg gcgggtacgt agctccccgc aatgaagccg
7441 agcggacgct tgccgatgtg tggcaggcgg tattgaacgc cgatcgcgtt ggggtaacgg
7501 atcatttctt cgagctgggc ggagactcaa tcaagtccat tcaagtatct tcgaggcttc
7561 atcaagccgg gtacaagctg gaaatacggg atttgttcaa atatccgact atctcacagc
7621 tcagcctgca tgtgaaaccg atcggacgta ccatcgatca aggtgaaata acgggtgaaa
7681 cggcgctgac gccgattcag cattggtttt tcgagagctc ctttgcggac ccgcatcatt
7741 tcgaccagtc ggtgatgctg taccggaagg aacgcttcga cgaagagacg gtacgtcagg
7801 tgctgcaaaa g[illegible]tggccgag catcatgacg ccttgcggat ggtgttccgc aagacggaac
7861 aagggtttag cgcgcggaac cgcgcgattc aggaaggcgg gctgttcacg ctggacgtgt
7921 tcgacttcaa ggatgcggag aataccgc[illegible]c aggctgtgga agcgaaggga acggacattc
7981 aagcgggcat cgatctggag aacggccccc tcgtgaaggc gggactgttc cggtgcgcgg
8041 acggcgatca tttgctgctc gcggttcatc atgccgtggt ggacggcgtg tcttggcgca
8101 ttttgatgga ggatttcgct ctgggttacg agcaggccgg caaaagcgag gaaattcgtt
8161 tcccggcgaa aacggatgcg taccgcactt ggtccgagca gctggccgct tacgcgcaaa
8221 gcccggagat ggcgaaggaa cgggcctatt ggcaggccgt ggaacaaatt gcggttccgg
8281 ccgtgccgaa ggatctggag gcggacgtta cgacgcagca ggacagcgaa tcgctgttcg
8341 tccgtttgac tcccgaagaa acggagctgc tgctgaagcg ggttcaccgg gcctacaaca
8401 ccgaaatgaa cgatattttg ataacggcgc tcggcatagc cgttcgcaag tggacgggac
8461 acgaaagggt gcggatcaat ctcgaaggac acggacgcga atcaatcgga acggatatcg
8521 acatcacgcg cacagtcggc tggtttacga ccaagtttcc ggtcgtcctg gagccggaaa
8581 ccgaccggga tttggcctat cagattaaac aggtcaagga aagcttgcgc cgcattccga
8641 acaaggggct tgggtacggc gtatgccgct atctctccaa atcggaggat ggctttgttt
8701 ggggcgcaga gccggaaatt aattttaact acctcggcca gttcgacgat gatgtcaacc
8761 aggacgagat cggcatatct tcttattcca gcggcagccc ggccagcgac cggcaggccc
8821 gcagctttgt gctggatatc aacggcatgg tgctggacgg cgctttatcg ctcgatctca
8881 gctacagccg gaagcagtat cgcaaggaaa cgatggaagc cttcgctcag cggcttgagc
8941 aaagtctccg agagctcatt acccactgcg caggcaaaga aaacaccgaa ttgacgccga
9001 gcgacgtgca attaaaggct tgaccatcgc ggaattggag caaatcgccc agcgctcggg
9061 ccatctcggg gaaatcgaaa atatttactc gcttacgccg atgcagaagg gcatgtggtt
9121 ccacagcgcg cttgaccgac aaacggccgc ttacttcgag cagacgcggt ttacgatgcg
9181 gggagcgctc gacgtccagc ttttcgagag gagctggatg gagcttgcga aacgtcatct
9241 ggtgctgcgg gcgaatttg tgaaaggacc ggcgggcgag ccgctgcaaa tcatataccg
9301 cgacaagccg gtcggctttg aatatgaaga gctgcttcat ttgcaggcgg acgagaaaca
9361 agcgtatttg gataaaaagg ccgaggatga caagcttcgc ggcttcgaca tggaaaatga
9421 cgcgctcgtt cgggttacga tcctgcgcac cgaagagcaa agctatcatg tgctgtggag
9481 tttccagcat attttgatgg acggctggtg cctgccgcag ctgacgcagg agctgtttga
```

```
 9541 gacgtactcg gccttggcat ccggcaagca gccagcggga aataagggat cggattatgg
 9601 cgcttatatc gaatggttgg agaaacagga cgatcaggcg gcatccggct attggacggc
 9661 attcctggca ggttatgaag ggcaaaccgt actcccggga caaaaggaac cggcgccaaa
 9721 cggcagattt acggctgatc acgtcaccgc cgagctgggc aaggacttga gcgagcggat
 9781 ggaccgggtg gcgaaacagc gcctggttac ggtcaatacg ctgctgcaag ccgcttgggg
 9841 cgtgatgctg caaaaatata acggaacaaa cgatgccgta ttcggcagcg tcgtggccgg
 9901 aagaccggcg gaaatcccgg gcatcgagtc catgatcggg ctgttcatca atacggtgcc
 9961 ggtccgcgtc acgagcgaag cggacactgt cttcgccgac ctgatggcga agctccaaga
10021 gcgggcgctg gagtccgggc gttatgatta ctatccgctg tatgaaattc aagcccgcag
10081 cgtgcaaaag cagaacctga tcaaccatat catcgctttc gagaactatc cagtggacga
10141 gcagatggag caggcgggcg accagcagca cggcgacctt acgatcgctg acgttcagat
10201 ggaggagcag acgaactata acttcaacgt gaccgtggtg ccgggagtcg agatcgaaat
10261 tcgattcgat tttaacgccg aagtgttcga taaagacagc atcgagttgc tcaaggggca
10321 tctcgtccat ctgctggagc aggtgacgga taacccggaa attaccgtgg gcgagctgga
10381 acttgtgacg gaggcggaaa aggccgacct tctcggacgt tttaacgaca ccaccacgga
10441 atttccgcgc gggaagacgc tcattcaatt gttcgaagag caggcggagc gcatcccgga
10501 tgcagccgcc atctccttga atgagcaaga gctgacctac cgcgagctga acgaacgcgt
10561 taaccgcctt gcccgtacct tgcgtagcca cgggatatcc aaaggccgtc tggtcgccat
10621 tatggctgag cgttccatcg aaatggtggt gggcatgctg gcggcataca aagccggagc
10681 ggcttacgta ccgattgacc cggaatatcc cgaggagcgt atccgtttct tgatcgagga
10741 ttcgggagcg caggtcatgc tgacgcaaag ccgcttgcgc gagcgcctgg cgggttcgga
10801 ctccgtgatc ttactggatg acgagtcctt ctatcacgag gacggcacaa atctaaatcc
10861 gggcaccgaa gcgacggatc tggcctgcgt catctatacg tcaggcacga cgggcaagcc
10921 gaaaggcaat cttgtttcgc accgcaacat cgtgcggatc gtgcgaaata cgaattatat
10981 cgacatcacc gaacgggatc atgtcctcca gctttcgagc tattcgttcg acggagcgat
11041 tttcgatatt ttcggcgctt tgaccaacgg ggcgcggctg gtgctggttc cccgcaagac
11101 tttgctggaa atcgtccggc tggcggatct catccagcgc gagcgcatct cggtcatgct
11161 gattacgacg gctttcttca acgtccttgt agatgtgaac gtcgattgcc tgcgggatgt
11221 ccgggcgatt ttgttcggag gagagcgagt gtcggtcgga catgtgcgta aagcgctcgc
11281 ccatatcgga ccgggcaggc tcaaccacct gtacggcccg tcggaaagca cggtttatac
11341 cacgtacctt ccggtcgact tcgtcgatga gtcggcggtt accgtaccca tcggacggcc
11401 gatcagcaat acgaaggtgt atatcgtcga cagccggaat aagcttctgc cgatcggcgt
11461 ggccggggag ctttgcgtcg gcggagaagg cttggtacgg ggctacaata accggccgga
11521 gctgacggcg gagaaatttg tggacaatcc gtttgtgctg ggagagcgca tgtaccggac
11581 gggggatttg gcgaaatggc tgccggacgg aacgatcgaa tacgtgggac ggacggacga
11641 ccaagtgaaa atccgcggct tccgcattga gctgggcgag atcgaagctc agcttcagaa
11701 agtggaggga attcggaaag caacggtatt cgcgagggaa gacgtctccg gcgagaagca
11761 gctttgcgcc tattatgaag cggaccgcga gcttccggcg gccgagctga agagcgtgct
11821 ttcccaggaa ctgccggcct atatgattcc ggcgtacctg atccagttgg agcggctccc
11881 gctgacgacg aacggcaagg tcgaccgccg atcactcccg gcgccggagg agagcttgca
11941 gccgggcgaa gaacgtactc cgcctcggac tccgctggaa gccagcttgg cgggaatttg
```

```
12001 gaaaagcgtg ctcggactgg agcacatcgg ggttcatgac aacttcttcg acctgggcgg
12061 ccattccctg cgggcgacga cactggtgag caaggtgcat caggagctga acgtcgagct
12121 gcctctgcgc gacgtattcc gctactcgac gatcgaagag atggctctcg ccatctcccg
12181 gatcggagag cagtcgttct cgtcgattcc gctggcgggc gcaagagcat attatccgct
12241 ttcctcagct cagaagcggc tgtttatcct gaatcagctg gaagggggccg atcagagcta
12301 caacatgccg ggcgtgctgc tgctggaagg atcgattgac cggagcctgc tggagaaggc
12361 tttccgcgga ctgatcgcac ggcacgaaac gctgcgaacc ggctttgaga tcgtacaagg
12421 cgaagctgta cagcgcattt acgagagcgt cgactttgcc gtcgagtacc gtcatgtgag
12481 cgaggaagaa gcgcccgaag tcgtgcaggc cttcatccgg cctttcgact tggcgaagcc
12541 tccgctgctg cgggcggagc tcgtagagct ggcagccgaa cgttatttgc tgatgttcga
12601 catgcaccat atcgtctccg acggggtttc gatggacgtg ttagtcgagg aactcgttcg
12661 tatgtacggc ggcgagtcat tagagccttt gcgcattcaa tacaaggact atgcggtatg
12721 gcagcagtcg gacgagcaaa aagtgcagtt gaaacgcgag gaagcttact ggctggaccg
12781 ttaccggggc gagctgccgg ttctggaaat gccgacggac tatccgcgtc ctgccgtgca
12841 gagctttgag ggacaaacgc tgacgtcctt cgtggacgag gcaacgaacg aaggcttgaa
12901 gcagctggcc gctcaaagag gaacgacgct gtatatggtg ctgcttgcgg catataccgt
12961 gcttttgcat aaatacacag gtcaggacga tttgatcgtc ggaacgtcga ttgcgggcag
13021 aacgcacgga gacacgcagc ctttgatcgg aatgttcgtc aatacgctgg cgctccgcaa
13081 ttatccggct tcggagaaga cctttctgtc gtatctggaa gaagtgaaag aaacgacctt
13141 aggcgcttac gagcatcaga attatccgtt cgaagagctc gtcgataaag tgcaggtcag
13201 ccgggatttg agccgcaacc cgctgtttga cacgatgttc tccctgcaaa acttggagga
13261 taaagagttt gagttggaag ggctgaaatt gtccccgtac cctagcgaat acggcacggc
13321 caagttcgac ctgagtgtgg atgttacgga agaaaatggc ggcctggagt gcagctttga
13381 attcgcaacg gctctttata aagaaagcac gatccggcgg ctgtcgactc atttcggaca
13441 tttgctcgta acgatcgtaa gccgtccgga tgcgaagatc gccgagctga acttgctgac
13501 ggcagaggaa aaagaacaaa ttctcggcgc gtttaacccg gcgcagcctg aagcggctcc
13561 tgtggcggcg ttccaccggc tgttcgagga acaggcggag cgcacgccgg aagcggaggc
13621 cgtcgtgtac gagaacgacc ggctgacgta tgcggagctg aacgagcggg cgaaccgctt
13681 ggcggctacg ctgcgcgcaa gcggcatcgg ccgggagacg atcgtcggca ttctcgccga
13741 gcgttcggtg gacttgctgg tggccgtgct ggccgtctgg aaagcgggcg gggcgtatgt
13801 gccgctcgac ccggattatc cggcggagcg cgtgcggttc atgctcgaag acagcggagc
13861 gaaggtactg ctgacgcaaa cggcgctgcg agagcgtgct gaagcctggc tcgacgaaga
13921 ggagctggcg ctggcggcgg tgctgtacct cgacgacgaa acttcgtaca gcgaggagcg
13981 gacgaatgcc gacagcttcc atgaagcccg tccggaggat ctggcgtacg tgatctatac
14041 gtcaggaacg acgggcaagc cgaagggcgt gatgatcgag caccgcagcc tggtgaacac
14101 ggcggcgggc taccggcggg aataccggtt ggatgagttc ccggtgcggc tgctgcagct
14161 cgccagcttc tcgttcgacg tgttcgtagg agatatcgcg cggacgctgt acaacggagg
14221 cacgatggtg attgtgccga aggacgaccg gatcgatccg tctcgtctgc accactggat
14281 ggagcgggag cgggtcacca tcttcgaatc gacgccggcg ctgatcgtgc cgttcctgga
14341 gtacgtgcac gagcagcggc tggatatgag ctggatggag ctgttgatca cgagctcgga
14401 cagctgcagc gtggcggatt accggacctt gcaggaacgc ttcggctcgt tgttccggat
```

```
14461 tatcaacgca tacggcgtga cggaagcggc gatcgactca agtttctacg acgagccgct
14521 gacgaagctg ccgcagacag gcaatgtgcc gatcggcaaa gcatggctga atgcgaaatt
14581 ctacatcgtg gacgcgcatc tgaacccggt gccggtcggg gtgctgggcg agctggttat
14641 cggcggagtt ggggtagcgc gcgggtactt gaaccgtccg gagctgacgg aagagaagtt
14701 cgtagacagc ccgttcgccg cgggcgagcg gctgtaccgc acgggagact tggcgcggtg
14761 gatggaagac ggcaacgtgg acttcatcgg ccggatcgac aaccaggcga aaatccgggg
14821 ctaccggatc gagacgggcg agatcgagtc gcagctgctg cgggtggaag gcgtgcgcga
14881 agcggtggtg ctggttcgaa gtgacgcgaa cgggcagaag gcgctatgcg cgtattgcac
14941 gccgggtacc ggagcggagc tggcggtgaa cgatttgcgc agcgcgctgg cgcaggagct
15001 gccgggctac atgatcccgt cgtacttcgt ggagctggag cgcctgcctc tgacgccgaa
15061 cggaaagatt gaccggaagg cgttgccagc gccggaaggg gaagcgggaa gcggaacgga
15121 gtacgtcgca ccgcgcaatg agctggaaac gaagctggcg gcgatttggc aggaggtgct
15181 ggggcttgcg aaggagattg gcgttcacga caacttcttc gacatcggcg gccactccct
15241 gcgggcgacg acgctggtca gcaagattca taaagagttg aacgtggatc tgccgctgcg
15301 cgacgtgttc cgccattcca cgatcgagag catggcggcc gccatttccc ggctggatga
15361 gcagacattc gttgccattc cggtggcgga tggccgagag gtgtacccgc aatcttttgc
15421 tcaaaaacgt ctctttatcc tgaatcaact ggaaggcgcg gagcttagct acaacatgcc
15481 ggaagcgatg ctgctggagg gtgctttaga ccgggcaagg ttcgaagaag cgttccgtaa
15541 gctcgtggcg cggcatgaaa tgctgcgcac cgggttcgaa atggtggatg gcgaagcatc
15601 gcagcggatt taccaggacg tgaattttgc cgtggagttc tatcgagcgg atgagcaaga
15661 ggccgaagag acggttcgcc gtttcgtccg tccgtttgac ttggcgaagc ctccgctgct
15721 gagggtaggc cttgttgagc tggctccgga acgccatatt ctaatgtacg acatgcatca
15781 tattatttcc gacggcgtct cgatggaaat ctttgttgaa gaattcgtcc gcttgtacgg
15841 cggcgagcaa ttggagcctc tgcgcattca gtacaaagac tacaccgttt ggcagcattc
15901 gcaggagcag aaggaacggc ttcagcgtca ggaggcgtac tggctggaca tgttccaagg
15961 cgagcttccg gtgctggaaa tgccgaccga ctatccgcgt ccggccgtgc agagctacga
16021 aggccaaacg ctggagtttt tctttgacgc ttcgaaaacc gacggcctga agcagctggc
16081 ctcggaaacg ggcacgacgc tgtttatggt gctgcttgcg gcgtataacg tccttctgca
16141 taaatattca ggtcaggaag acgtgatcgt cggtacgccg attgccggaa ggaatcatgg
16201 agatgtgcag cagttgatcg ggatttctt aaacacgctg gctatccgca gttatccggc
16261 ttcggagaag acattcctgt catacctgaa cgaagtcaaa gaaacgaccc tccacgcctt
16321 cgagcatcaa aactatccgt tcgaagaatt ggtggacaag gtgcaagtca cccgtgattt
16381 aagccgcaat ccgctcttcg acacgctgtt tacgatgcag aatacggaga acgaggaatt
16441 tgagctggaa gggcttcgcc tgattcctta tccgagcgca ctggataccg cgaagtttga
16501 tatcagcttg gatgtgagcg aggagaacgg cggcttggat tacagcttcg aatatgcgac
16561 ggctctctac aaaagggaga cgatcgaacg gctggcaaaa cattacgagc agctgctcgt
16621 aacgatcgta agccgtccgg atgcgaagat cgccgagctg aacttgctga cggcagagga
16681 aaaagaacaa attctcggcg cgttccaccc ggcgcagccg gaagcggctc ctgcggccgc
16741 gttccactgg ctgttcgagg aacaggcgga gcgcacgccg gaagcggcgg ccgtcgtgta
16801 cgagaacgac cggctgacgt atgcggagct gaacgagcgg gcgaaccgct tggcggctac
16861 gctgcgtgca agcggcatcg gccgggagac gatcgtcggc attctcgccg agcgttcggt
```

16921 ggacttgctg gtggccgtgc tggccgtctg gaaagcgggc ggggcgtatg tgccgctcga
16981 cccggattac ccggcggagc gcgtgcggtt catgctcgaa gacagcggag cgaaggtgct
17041 gctgacgcaa acgccgctgc gagagcgtgc cgaagcctgg ctcggcgaag aggagctggc
17101 gctggcggcg gtgctgtacc tcgacgacga agcgtcgtac agcgaggagc ggacgaatgc
17161 gccgctggaa gcggatgatc acttgcacga atgccgtccg gaggacctgg cgtacgtgat
17221 ctatacgtcg ggaacgacgg gcaagccgaa gggcgtgatg atcgagcacc gtagcctggt
17281 gaacacggcg gcgggctacc ggcgggaata ccggttggat cagttcccgg tgcggttgct
17341 gcagctcgca agcttctcgt tcgacgtgtt cgtgggagat atcgcgcgga cgctgtacaa
17401 cggaggcacg atggtgattg tgccgaagga cgaccggatc gatccgtctc gtctgcacca
17461 ctggattgag cgggagcggg tcaccatctt cgaatcgacg ccggcgctga tcgtgccgtt
17521 cctggagtac gtgcacgagc agcaactgga tatgagccgg ttggagctgt tgatcacgag
17581 ctcggacagc tgcagcgtgg cggattaccg gaccttgcag gaacgcttcg gctcgttgtt
17641 ccggatcatc aacgcatacg gcgtgacgga agcggcgatc gactcaagct tctacgacga
17701 ggagctggcg aagctgccgc agacaggcaa tgtgccgatc ggcaaagcgt ggctgaatgc
17761 gaaattctac atcgtggatg cgcatctgaa cccggtgccg gtcggggtgc tgggcgagct
17821 ggtcatcggc ggagtcggag tggcgcgcgg gtacttgaac cgtccggagc tgacggaaga
17881 gaagttcgta gacagcccgt tcgccgcggg cgagcggctg taccgcacgg gagacttggc
17941 gcggtggatg gaggacggga acgtggactt catcggccgg atcgacaacc aggcgaaaat
18001 ccggggctac cggatcgaga cgggcgagat cgagtcgcag ctgctgcggg tggaaggcgt
18061 gcgcgaagcg gtggtgctgg ttcgaagtga cgcgaacggg cagaaggcgc tatgcgcgta
18121 ttacacgccg gataccggag cggagctggc ggtgaacgat ttgcgcagcg tgctggcgca
18181 ggagctgccg gggtacatga tcccgtcgta tttcgtggag ctggagcgcc tgcctctgac
18241 gccgaacgga aagattgacc ggaaggcgct gccagcgccg gaaggggaag cgggaagtgg
18301 aacggagtac gtcgcaccgc gcaatgagct ggaaacgaag ctggcggcga tttggcagga
18361 ggtgctgggg cttgcgaagg aggttggcgt tcacgacaac ttcttcgaca tcggcggcca
18421 ctccctgcga gcgacgacgc tggcgggcaa ggtatttaag gaattaaacg tcaacctgcc
18481 gttgcgcgac gtattccgtc actcgacgat tgcggcgatg gccgaggcga tcgcccggat
18541 ggaacggcgg gagcatgagg ccattcctca agcggaggag agagagtact accctctgtc
18601 ctccgcgcag aaacggctgt tcattcagca cacgctggat ggagcggatc agctttacaa
18661 catgccggag ctggtgcagg tggaaggcga gtttgatata gaccggttgg aagcggcctt
18721 gcggaaattg ataacacggc atgaatcgct gcgcaccggt tttgaaatcg tgaaggccga
18781 agcggttcag cggatttacc cgcaggtcga ttttgctatc gagcatcatc aagcggataa
18841 agaggatgcg gctcaaatcg agcagatcgt ccgcagcttc gttcgtccgt ttgatctcgg
18901 caagccgccg ctgctgcgcg ccggggtcat cgagctggag ccgaacctgc atattctcct
18961 tttcgacatg caccatatgg tgtccgacgg cgtatcgatg gcgattgtga tcgatgagtt
19021 ctcgagtttc tacgccgggg aagaactgcc gccactgcgc attcaataca aggattatgc
19081 cgtttggcag cagtcgaagg cccaccgaga gcggatcggg cggcaggaag tgtactggct
19141 gcaaaccttc gaaggcgagc tgccgacggc ggacctgccg atggactacg aacggtctgc
19201 ggctcgcagc tacgaaggcg cgcatctgga gttcgacgtc gaagcttcgc tctctgcgcg
19261 gctgcgcgaa ttggcggccg agcgtgaaag cacactgttc atggtgctgc ttgcggctta
19321 taccgtgctg ctgtccaagt acagcgggca ggaggacttg gtcgtgggca ccccggtggc

```
19381 gggaagaacg aacgccgatt tggaaccggt catagggatg tttgtcaata cgctggcgat
19441 ccgcaatcgt ccgtcgggcg acaaaacgtt cttgtcctac ctggaagaag tgaaggaaac
19501 ggctttgggt gctttcgaga accaagatta tccattcgag gagctcgtgg agcgtttgaa
19561 tgtgaagcgg gagccgggcc gcttcccgct gttcgatgcc gttttcgact tgcaaaatat
19621 cgaagaacga gacgccgagc tggaaggggt cagcctgaag acttacgagc ttgaccattt
19681 ggaagaagcg aagttcgatc tgacgctgtt tatgtatgaa aacaacgggg cgctgagcgg
19741 gggcttcttc tacgccacca agctgttcaa agaagcgatg atccgcacct tgaccgagga
19801 ttacctgcgg gtactgtctc aaattgcgga aaatccacaa ctcgagctaa gccggattga
19861 atgtcataaa ccgacggcag gcgcaaagag tgccgtcgat acgatcgaat tcgcgttcta
19921 atcctacaag cgcgcctccg ccgtcaagcg aaagcggcgc gcatcccaag gaagacggcg
19981 aacaacgtcc aagcccgtaa acgcgcaggc cgaaagcaag ctttttcgga cctgcgcctg
20041 ggcatgcggt cacttcattt ttaggggagg tacaaatgaa atctttgttt gaaaaggaag
20101 aacagtactg gagcggcaag tttgacgccg atgacagcct gagcttcctt ccctacagtc
20161 aatcctccaa attatccgcc gacggggaag ctgcggccga gccgggcttg cttcaccgta
20221 ctctgccgag cgaactctcg gagagaatca tttgcctcgc caacggttcg gatttggctt
20281 tgtacatgat tgttttggca ggagtaaaaa gcctgctgtt caagtatacc gggcgggacc
20341 aagtgctggt cggcatgcct tcttatagcg cagaccccga cgggactccg ccgccgcatg
20401 acatcttggt gatcaagacg tccgtaagcc gccagactac gctgaaaacg ctgctcgggg
20461 gcatcaaagc ctccatcggc gaggcgctgg agcatcagca cctgcctttt cggaaaatgg
20521 tggagccgct ccatctggac tatacggggg acggcctccc ggtcgtcaac accgtcgtat
20581 ccttcgctcc gattcatccg gaaccgctgg gtaaccgggt ggcggccgat acgtttttc
20641 gcttcgaccg ccaaaaccac tccatcgagc tggaaataag ctttgacggg cagcggtacg
20701 agcgcgcatt tgtggaacag gcggccgacc atcttgttag gctgctgtcc gtgcttttat
20761 ttcagccgga tctggagctc ggacaagccg atgtgctgtc cccagacgag agggagacgc
20821 tgctgaagcg atttaatgac accgaaagcg ggttcgagcg ggggaaaacg atttacggct
20881 tgttcgaaga acaggcggag ctttacccgg acaacgtggc cgccgtcatg aacgagcggc
20941 agctgaccta ccgcgagctg aacgagcgat ccaaccgcct tgcgcggaag ctgcggaaga
21001 cgggagtaga agcggaccag ctggtagcga ttctggccga acgctcgctc gatatggttg
21061 tcggcattct ggcgattctc aaagcgggcg gagcatacgt gcctgtcgat cccgactacc
21121 cggaggagcg catccgcttc atgatcgagg attcgggcgc gccgttattg ctgattcaaa
21181 agcatctgca cgagaagacc gacttcgcag gaacgcgcct cgaattggac gatttcgttt
21241 ggggcgacag aggggcggac tccgccgatg cgctggacgc ttcgaacctg gagccgattt
21301 ccgggccggg caacctggct tatgtcatct acacgtcggg aacgaccggc agaccgaaag
21361 gaacgctgat cgagcataag aacgtcgtgc gcctcctgtt caacgacaag aacctgttcg
21421 acttcgggcc gtccgacacg tggacgctgt tccactcgtt ctgcttcgat ttctccgtct
21481 gggaaatgta cggagcgctg ctgtacggag gcaagctggt catcgtaccg ccgctcacgg
21541 cgaaaaatcc agccgatttc ctggcgctgc tgggccgcga acaggtcacg attttgaacc
21601 agacgccaac gtacttctac cagctgctgc gtaaggtctt ggcggaccat ccgtacgatc
21661 tgcggattcg caacgtcatc ttcgggggcg aagcgctgag tccgctgctg ctcaagggct
21721 tcaagacgaa gtacccggag acgaagctga tcaatatgta cggcattacc gagacgacgg
21781 ttcacgttac gtataaggaa atcacgtggg tcgaaatgga ggcggcgaag agcaatatcg
```

```
21841 gcaagccgat cccgacgctg agcgtgtacg tccttgatga aaaccgccgc cctgtgccga
21901 tcggcgtagc gggcgaaatg tacgtggccg gggaaggcct tgcgagagga tacctgaacc
21961 gcccggatct gacggcggag aagttcgtcg attccccgtt tgcggagggg gagaaactgt
22021 accgctcggg cgacttggcg gcttggctgc cggacggcaa catcgaatac ctggccgga
22081 tcgatcatca ggtgaaaatc cgtgggtacc ggatcgagct ggacgaaatc gagacgcagc
22141 tgctgaagat cgccgccgtg caagaagcca aggtgctcga ccgcgacgac gcgaacggcc
22201 ataagcagct tgtcgcttat tacgtcgcgg aaacgaggct ggcggcgcat gaactcaagg
22261 aggagctcgc caagcagctt ccggggtata tgattccttc gcacctcgtg cagctttcgc
22321 ggatgccgct gaccccgaac gggaaaatcg accgcaaagc gctgcccgcg ccggaggaag
22381 ccgcggccgg aggagcggaa tatgtcgcgc cgagaacgct gctcgaaatg aagatcgccc
22441 gcgtttggca ggatacgctt ggcgttccgc aggtcggcgt aaaggataac tttttgagt
22501 tgggtggcaa ttcgttaagt ctgatgaggc tcgttcaagc cgtttacgat gaaacggaca
22561 tcgagatacc gctgaaccgc caattccaca atctaacagt tgaagctatg gctttcggca
22621 agggggatct gggcctggat aaagggggag actccttcat taagctgaat aaagcaggag
22681 atctgaacgt gttctgtttc cctccgggca gcggcttcgg catcggttac cgagagctcg
22741 caagcaggct cgacggccgg ttcgtgctct acggcattga ttttatcgac gatgccgtcg
22801 attacgaggc catgctgaac cgttatgtgg acgagatcgt ccgcatccag tcggaaggac
22861 cttacgtgct gctcggctac tgcttcggag gcaacctgac gttcgaggta gccaaaacga
22921 tggagaaaag agggtattcc gtaacggacg tgctcatggt ggactcgtgg attaaggaca
22981 cgctgacgcc ttccgaaacg tcggagaaag agcttgaaga aacgcttgcc gatttcgacg
23041 aagaagagaa tgaattaatg agcaacccgc tcgtgcggga gcgggttcat cggaaggtca
23101 aagcgacctt ggcgtacgaa gcgcagctta ttaattccgg cacgatcccg gcccggattt
23161 acgaactgat tgcgaaggac agcgaagcgt tccgcctgga gcaccaatcg ccttcctggc
23221 gggggcaac gacgcaagct tacgccgatt accggctgga gggcgcgcac gaggaattgc
23281 tggaactcgc gcgcgtggaa gaaacggccg tcgtcatccg ggacatctta gagcaagtca
23341 agcggcagat cgaagcggag gccggggtac tgcatggaag ctgaccgcca gccgtctgtt
23401 ttagaacaag gcacagcggc gccttccaag cgccaaaccg cttacgccgt ctggaaagcg
23461 ttccgctggc tgatgtccta tgtaagccgt cacaaaggct ggatgatcgt cggtatcttg
23521 tccgcgattg ccgccgccgt tattgagata tggatgggaa gtttgatcga gcagctgacc
23581 acccaggccg agaaggggc ggggccaatc gttctgcaaa tcgtgtacac ggtctttgtg
23641 gtcatcttga tcggtgtgcc ggcgaagttt ttcatgagct tcggcgtgga gcgaagcagc
23701 gcctctgcag tgcaggacat ccgcaaccat gtcatgcgtc atatcggcaa actcccggtc
23761 tcctatttgg aaaagcagca ctccggcgac gtattgtcgc ggatcaacaa cgacctgcag
23821 ctcatccagc agtttatgat tcgggacctc gcccagtggt tttatcatcc gctggtgttc
23881 atcggctgtt tcgcttattt gatctacctc caatgggagc tgatgctgta cagcctgctg
23941 ttatttcccg tagcgctgct ggtttcccaa tggatcggca agcagttgga gcggttgacg
24001 gaggaagccc aggcgaacat gggccgaatg aacgtcaacc tccaggatac gcttggggt
24061 atgcctattg taaaaagcta cctgctatcc ggcatgttat ctcgctccta ccaagtgctg
24121 ctgcaattga cggcccaaaa aaagctggcc gtgaaaaagc gggaagcctg ggtcaacccg
24181 ttgctttcca cgctgatgat cagcccgatc attttcgccg tcagttacgg aagctatttg
24241 atctacaaag ggcagctagg cgcgggagag ctgatcgcct tcctgtactt gctgaatctg
```

```
24301 tgtctggagc cgctggagca tattcccgag ctcatcacgc ggacgttcga aatggccggt
24361 gccctgagaa gggtctccga aatcgtcgag cagccgaccg aaacggaaaa tggccgttgg
24421 cttccgaaag cgagcgccgc ccccatcgag tttcagaacg taaccttcgg gtatgaggag
24481 agctccccga tcctgcggaa tgttagcttc tcggtgccgg aagggaagac gatcgcgctt
24541 gtcggagcga gcggcggagg gaagagcacg gtgtttaagc ttgtatgcgg cttttatccg
24601 cttccggagg accagggcga gatccgcgtg ttcggcagcc tgatccacgg cgccgatccg
24661 gagcagcttc ggtcgcattt ttccgtggta acccaggatt catatttgtt tagcggcacg
24721 atcgctgaaa atatcggcta cgggcgggaa gtagcgtcga tggacgagat tatcgaagcc
24781 gccaaagccg ctcaggcgca ttccttcatt atgcagcttc ccgacggcta ccagacgtat
24841 gtcggagagc gcggaggcct tttgtccggc gggcagcgcc agcgcattgc aatggcccgg
24901 gcttttctta aggatgctcc cgttctgctg ctggacgagc caacgtcggc tcttgatccg
24961 gagtcggaaa gcgcggttca ggaagcgctg ggcgtattga tgaagcagag aacgacgatg
25021 gttattgccc accggctttc tacagtacaa aacgccgacg aaatttgggt catggaacag
25081 gggagtattg ttgaaaaggg cactcatgaa caattgctga agatgaaggg actgtacgcc
25141 cagtcgtact atcaggaatt taccgagtct gccgaacgca gggaggtggc gtacacatga
25201 aaaagggcgg atggctctcg caagtgaaag aactcggcta cttgctgatg tttatgaacc
25261 gcaagcgcaa aacccagtac gccattggcc tggccgtgac ggcgctcacc cagacattgt
25321 tcttgatcgc cttcagcttg gtcgtacata acttggttga tttcgccgtg tcccgagata
25381 cgtccctgat ggtggaagcc tttattattt tgggcgcggc cttgttcctg gaaaatgtga
25441 tttctccctg gttcatttac ttataccagc gcagtgtcga gctgactgtg ctgaatatcc
25501 gcgaacgtct ttatgacaag ctgtgccggg tgcggccgag attcctggag cagacgcacc
25561 atggggattt gctgtcgcgg gtgaacaacg acgtaacgac cgttgagttt acattctcgc
25621 aggtttactt cgttttgctg cttcaagttg tcttttgcat cggctccatt gtctccatgg
25681 tgttgatcga ttggcggttt gccggcgtat ccttcgtcat tctgctgctg tcctccgtgg
25741 tcagcctgaa atttgcgcgg gatattcgcg ccttgtccga acaaggcttg caaacgctcg
25801 gtaaaatgac cgaaaaattc aaagatttta tgggcggcat tcaaattgtg aagctgttcc
25861 gcatccgcac gatttacggc cagtacgagg cgttgaacga acaaatgacg cagacgcttc
25921 ggcaaaccgc gcagaagaac ggcatgcagg ctgcggtgaa tcattttatc agctacgtca
25981 cgttctgcgg catcatcgtc atcggcagtc ttctttatgc ctacggactg atgggaatgg
26041 gaagcgtcgc tgccctggca gttctgcaag tgaatctgac gcacgccttg ttgaacttgg
26101 gtatggttct gtcgatgacc caaaattcga tcgcgggcgc tcaccggatt caagaggtac
26161 taagagagga agaggaaccg gagcgtttgg gatctcctca cagcgagctc gtgtcggagg
26221 ctgcggtgga gtttcgcgat gtggaatttt cctatcaggc ggataaaaag gtgcttgtcg
26281 acatctccat gcaggtgttt cccggccagg tcgccgctat cgtaggggcc agcggcagcg
26341 gcaaaagtac gctgatcaag ctgctgctcg gcttttatcc tgtggacagc ggagaaatcc
26401 tgctccaagg caaaccgttc ggccattaca cgctggacga aatccgcagg cagattgcat
26461 acgttccgca ggagccgttc ttgtttaccg gcacgattga ggaaaacatc cgctacggca
26521 acccggatgc aacggatgaa gaagtgatcg aagcggccaa agcggcgtac gctcaccatt
26581 tcattcagga acttcctgaa cagtataaaa cgccggtggg agagagagga gcgtcgttgt
26641 cgggcggaca aaggcagcga attgcgatcg cccgggcgat tctcaaaaac gccccgattc
26701 tgctgctgga cgaagcgact tccgcgctgg ataacgaatc ccagcattgg gtacagcagg
```

```
26761 ccctgaacgt attgatgaag gggcgtacca ccattctgat cgcccaccgt ctcagcaccg
26821 tggaacatgc ggatttgatt accgtcatga accaagggac ggtcgtcgag cggggccgcc
26881 atcaggacct gctggcgctc ggggttatt acgcccggct gtacggctag acccgcttta
26941 cggcgggcgg cttcctgctt gcgaatgcga cagagtcttt tagtaatcta aggtatgact
27001 ttataaccgg tagttaccag aaaacgtaga cacgaaaaag gaggaagagt ccgaagtagc
27061 cgttccggga atggcctgcg cggttcttct aagagtgact aagtgacatc taggtacaaa
27121 gcctatctat aaagaatctc taaaccggga gtgtgtcgat gtgagagaga ataccaacga
27181 gcaatatgga ttaacgcaag cccagcgccg aatatggttc atggaaatta tgaatccggg
27241 aacgtccatc acgatgcttt ccgcgaccta ccagattacg ggcgagatca acacacagct
27301 tctggagcaa gcggtggcag agatcgtcaa aacctatgac gttttccgaa tccgcattag
27361 cggggatttg caaaatccaa cgcagtggtt cgaagagccg gagaatgtcc aagctaggat
27421 aagccgcctt gaaataggca caaccgaaca attctatgct tgggtgaaag aagtaagcga
27481 aaaaccggcc agcgtgttcg acgaacacct ctaccaattt acgattatcc attttgcgaa
27541 cggccaagta tggctcaatt tgacggtaaa tcatattatc gccgacggct tgtccgtcaa
27601 tgctttgctg catgcggtga tggaaaaata cctggaactg cgcaaaggca tctccagcag
27661 ttaccaggcc ccttcctatc tggattatat ttccgcggag cgtgaatatg agcaatcgca
27721 gcgttatcaa aaaggcaagg aatactggct gacgaagtac aacactttgc ctgaaacgac
27781 cggcattaaa tcgtatccgc cattctcgat cggcagcgaa tccaataaac tggccatcac
27841 tttggacggt tcccggtatg aacgcattct ggccttcagc gaacaatatc aggtcagctt
27901 atatacgtta tttctgtccg ccatgtacgc cttattgtac aagctgaccg acagcaccga
27961 tgttccggtc ggcacggttt tcgccaatcg caccagcaag aaggaaaaag aaacgatcgg
28021 catgttcgtc agcaccgtgg ctacgcggat tcgtctgaat ccagacgggc acgtgctttc
28081 cttgatccaa acggtttcca aggaaaatac ggcggatctg cggtatcaaa aataccctta
28141 taaccaattg atccaggatt tacgcgaaca acacggccgc aacgatcttt cggggctgtt
28201 ccgcacgtct ctggaatatc tgcctttgaa aatcgtggag tacgaagaaa tcaaggtacg
28261 cctggaggct cacttcgcta ggcatgagat ggacgatttg ctgctgcgct tcgaccatat
28321 gctgaatgaa ggccatgtaa ttctccatgc ttcctatcgt accggcctgt tcgagacggc
28381 cgagattgat cggattatgg aacagtatgt aaccgttctg gaccagtttc ttcagactcc
28441 cgaactgccg atacgcgaga tttctctgct gagcgatgag gagagacacc gcattctcaa
28501 cgtttttaac ccgccggtgg cagggctgag cgagggagaa gcgtttcatc ggtatgttga
28561 aaagtttgcc cgggaaattc cggatcatcc ggcagtcgtc tacatggaca aaaagctgac
28621 ctacggcgaa ttgaacgaac gcgccgagcg gctggcttct ctccttcgcg aacagggcgt
28681 gggaagggag acgattacgg ggatctgggc ggagcgttcg gtggaactgc tggtcggggt
28741 gctcgccgtt tggaaagccg gcggagccta tgtaccgctg gaccctgatt atccggcgga
28801 gcggattgag tacatgctca gcgatagcga tgcatcggtg ctgcttacgc agcgtcatct
28861 gttggagcgg accagaggtt ggttggccga tgaccggctg aagcttcaag ctgtctatgc
28921 catggacgat gaacagattt ataacgggga tgccttagcc gtggaatttg aatctgccgg
28981 cagcgccccg caagacttgg cttatgtgat ttacacctcg ggtacgacgg gacgcccgaa
29041 aggcgtcatg atcgagcacg gtagtctcgt gaatacggcg gatgcgtacc gtcgcgagta
29101 ccagttggat cagtttccgg tgcggctgct gcagctggcc agcttctcgt ttgacgtgtt
29161 cgtcggagac atcgcgcgga cgctgtataa cggaggcacg atggtgattg tgccgaagga
```

```
29221 tgaccggatt gatccgaacc tcttatacgg ctggattcgg gaccaaaaca ttacggtatt
29281 cgaatcgacg cctgcgctca tcctaccgtt catgcagcat atttatgaag aagggctgga
29341 cgttagctcc atgcagttgc tgattaccag ctcggatgct tgcagtgtca ccgattaccg
29401 attgctgcag gaaagattcg gcggacaatt ccgcatcatc aacagctatg gcgttaccga
29461 agcggccatt gacagcagct tttacggtga gccgctggat aagctgccgc cgtcgggtca
29521 tgtgccgatc ggcaaagctt ggctgaacgc ccggttttac attgtcgatg ccgcgttaaa
29581 gccggttcct gtaggggttc cgggcgagct tgtcatcggc ggcgccgggg tggcgcgcgg
29641 gtactggaac cgtccggacc taaccgccga gaagtttgcg gacagcccgt ttgtaccggg
29701 cgaacgtctg taccggacag gcgatttggc ccgctggctg gaagacggca acgtcgactt
29761 catcggccgg atcgactatc aggtgaaaat tcgcgggttc cggatcgaac tcggcgaaat
29821 tgaaacggcc ctgctgcgtt tccctggcgt caagcaggct gtggtgaccg accgtacgga
29881 tgagcagggg caaaagtatt tgtgcggcta cgtggcggcg gatgcttcct tgcagctgag
29941 cgatctgctg tcccaattga agcaagagct gccggcccat atggtcccgg cccggctggt
30001 gtctcttgat aagcttccgc tcacgccgaa cggcaaaatt gaccgtaaag cgctgcctga
30061 accgaccgga gaggtagaag caggccgtga gtatgtggct cctcgcacaa cgctggaaac
30121 aagacttgct ctcatttggc agcaggtgct gggtattgcg cgagttggag tccaagacga
30181 tttctttgac ttgggcggtc attccttgcg ggcctccacg cttgtttcca agattcggaa
30241 agagctgcaa gtcgaggttc cgctgcggga agttttccgc tacaccacga tcgaacagct
30301 ggcccaaaga atcggcggtt taaggcagca ggagacgtat gagattacaa aggcggctga
30361 ggccgagtac tatccggttt catccgagca aaagcgtctg tacgtcctgc gccagcttga
30421 cggggccgag cgcagctaca atatgtcggc ggcgcttctt ctcgaaggca agctggaccg
30481 cacgcgcgtc gagcacgcgt tccgggcgct gattcagcgc catgagacgc tgcgtaccgg
30541 gatcgagcag gttcaaggcg agcttgtcca gcgcatctat gacgaggtgg agtttgctgt
30601 cgattatttc caagcgagcg agcgggaagt ggagcaagtg gtggaagctt actatcaccc
30661 gtttgatctg accaagccgc cacttctccg catcggcctg atcgaagtcg ccgcggatcg
30721 ccacattctg ctgttcgata tgcaccatat cgtctcggac ggcatttcga cagcgctgct
30781 cttcgacgag ttcagccgcc tgtatcgggg cgaggagctg gccccgctgc gtattcaata
30841 caaagattat gccgtttggc agcattccga agcttacggg cagctgctcc agccgcagaa
30901 ggagtattgg ctggaacagc tgtcgggcga gctgccggtc ttggagctac cgacggactt
30961 cccgcggcct gcggtgcaaa gctttgacgg ccgaaccgtg aagttttata tcgagaaaga
31021 tcggacggag aagctgaaag agctggcggg acggacaggg acgaccctgt acatggtgct
31081 gctgtcggct tacaccatcc ttatgcataa atattcgggt caggaagatc tgatcgtagg
31141 cacgccgatt gccggaagaa cgcaggaaga agtgcagccg atcgtaggga tgtttatcaa
31201 cacgctggcc attcgcagcc gtccggagcg ttccaagccg tacctttcgt acctggaaga
31261 aatcaaggac atcacactcg ggctttcgaa acaccaaaat tatttgttcg aagacttggt
31321 ggaaagtctt cacattccgc gcgcgaccgg ccggaatccg ctctttgaca cgttcttctc
31381 cctacaaaat acggagaacg agcaaattgt catcgagggg ctggagcaat cgttttatcc
31441 actggaaaac caaacatcca agttcgagct gctcctggat atttcggagc tggacggtca
31501 gctcgaatgc cggttggagt acgctacggc tttgtataaa caggagaccg cggagcggtt
31561 cgccagacat tatgacaagc tgctcgaaac catcgcagca gggccggacg ggatattgc
31621 ctcgctggaa atgctcacgg aggaggaaat ccgcgagctg gtgcgcggtt tcaacgattc
```

31681 ggaggcggac tacccgcggc agcagacgat tcacggcttg ttcgaagagc aggcggagct
31741 ttacccggac aacgtggccg ccgtcatgaa cgagcggcag ctgacctacc gcgagctgaa
31801 cgagcgatcc aaccgccttg cgcggaagct gcggaagacg ggagtagaag cggaccagct
31861 ggtagcgatt ctggccgaac gctcgctcga tatggttgtc ggcattctgg cgattctcaa
31921 agcgggcgga gcatacgtgc ctgtcgatcc cgactacccg gaggagcgca tccgcttcat
31981 gatcgaggat tcgggcgcgc cgttattgct gattcaaaag catctgcacg agaagaccga
32041 cttcgcagga acgcgcctcg aattggacga tttcgtttgg ggcgacagag gggcggactc
32101 cgccgatgcg ctggacgctt cgaacttgga gccgatttcc gggccgggca acctggctta
32161 tgtcatctac acgtcgggaa cgaccggcag accgaaagga acgctgatcg agcataagaa
32221 cgtcgtgcgc ctcctgttca acgacaagaa cctgttcgac ttcgggccgt ccgacacgtg
32281 gacgctgttc cactcgttct gcttcgattt ctccgtctgg gaaatgtacg gagcgctgct
32341 gtatggaggc aagctggtca tcgtaccgcc actcacggcg aaaaatccgg ccgatttcct
32401 ggcgctgctg ggccgcgaac aggtcacgat tttgaaccag acgccaacgt acttctacca
32461 gctgctacgt gaggtcttgg cggaccatcc gtacgatctg cggattcgca acgtcatctt
32521 cgggggcgaa gcgctcagtc cgctgctgct caagggcttc aagacgaagt acccggagac
32581 gaagctgatt aatatgtacg gcattaccga gacgacggtt cacgttacgt ataaggaaat
32641 tacgtgggtc gagattgagg cggcgaagag caatatcggc aagccgatcc cgacgctgag
32701 cgtatacgtc cttgatgaaa accgccgccc tgtgccgatc ggcgtagcgg gcgaaatgta
32761 cgtggccggg gaaggccttg cgagaggata cctgaaccgc ccggatctga cggcggagaa
32821 gttcgtcgat tccccgtttg cggaggggga gaaactgtac cgctcgggcg acttggcggc
32881 ttggctgccg gacggcaaca tcgaatacct gggccggatc gatcatcagg tgaaaatccg
32941 cggctaccgg atcgagctgg acgaaatcga gacgcagctg ctgaacgtcg agggcgtgga
33001 agaagcggtg gtgctcgccc gtcaggacgg cgggggcgag aaggcgcttg tcgcctactt
33061 tgtggcggac cggacgctga cggccagcga aatgagaacc tcgctggccc aggaaatgcc
33121 ggggtacatg atcccgtcgt acttcgtgca gctggagcgc atgccgctga cgtccaacgg
33181 aaaagtggac cgcaaagcgc tgccggagcc gcaaggcggc ttgcaaacgg gcgtcgaata
33241 tgtagcgccg cgtaaccgga cggagtccca gcttgtgaag atctgggagg aagtgctagg
33301 ctactccggc attggagtcc tggacaattt cttcgagctc ggaggccact ccttgcgggc
33361 gacgaacctt gtcagcaaga ttcagaagga aatgaacgtc gaacttccgc tgcgcgatgt
33421 gttccgctat acgacgatag agtcgatggc cggggctatt gccagcttgg aagaaacgcg
33481 gcacagctcg attccgaaag cggaagagaa ggcgtactat ccggtttcct ccgcgcaaaa
33541 aaggctgtac gtcctgcacc agctggatgg ctcggagctg aattacaacc tcccaagtgc
33601 cttgcagttg gaaggggctt tgaacgaggc caaagtggaa aaggcgctga caactttggt
33661 ggcccggcat gctatgctgc gcaccggttt tgaaatcgtg aatggggagc cgattcagcg
33721 tattcatccg ctcgtagctt tcaaggtcga gaagcttcaa gcaagcgaag atcaggttgc
33781 ggccattctt gaaggcttca ttcagccttt tgacttgacc cagccgcctt tgctgcgcgc
33841 cctgctgatc gaactggaga aagagaaatt cctgctcgcg ctggatattc atcatattgg
33901 ttccgacggc ctttccatgg acgtgctgct gcgcgaattc gtgcggcttt acaatgggga
33961 agaattgccg gagctgcgga ttcaatataa ggattacgcc gtttggcagc aatccgagga
34021 gcagcgccag cgcatcaaac ggcaggagga atactggcgc ggggtattca gctccgagct
34081 tccggttctt gagctgcctc tcgacttctc ccgcccggcc gtccagcagt ttgacggtca

```
34141 aacgctcacg tttacgctgg atgcggagaa aagcgaagct ctcaaacggc ttgccggcga
34201 ttcgggggcg acgctgtata tgcttctgct ggctgcgtac tccgtattgc ttcataaata
34261 cgcgggacaa gaagatatag tggtcggcac tccgattgcc gcccgttctc acgccgacct
34321 gcagccgatt atcggcatgt tcgtcaatac gctggccctt cgcttgggtc cggcggcgga
34381 gcggacgttc ctggattact tgcaggaagt gaaggaaacg acgcttggag cctacgagca
34441 ccaggactat ccgttcgagg agctggtgga agctcttcag gtgagccggg atttaagccg
34501 gaatcctctg tttgacacca tgttttcttt gcaaaaacac gaaagcttgg atttaaccct
34561 ggaaggcttg caatggtcgc tgttcgacat cgaggaaaag acggcaaagt ttgatcttag
34621 ccttgatatc gtggaagccg ataacgagct ggtttgcaag atcgagtacg ctacctcgtt
34681 gtttagacag gaaacaatgg tacggctggc gggtcattac gagcagctat tggcgtcgat
34741 cctggctcag ccgggtgcgc ggatttcgga tttggatatc ttgacggaca gcgaaaagca
34801 tgatttgctg atcgggtttg acgtgtcgtc ttcggctctt gcgaagcaac acgccgcaga
34861 gggtacaggt ttggaagcgg atgagtcgtg gagagagaga acgttccacg agctgtttga
34921 ggagcaggcg gaacgcactc cgggagcgct ggcggttgtc tacgaagaca gcaagctgac
34981 gtatgcggag ctgaacgcca aagcgaatcg tctggcgtat gcactgcggg cgcgcggggt
35041 gaagccggag caggtggtcg gcattctggc tggccgttcg gcggagctgt tgatcggggt
35101 gctcgccgta tggaaagcgg gcggcgctta tgtgccgctc gacccggact atccggcgga
35161 gcggatcgag tatatgctcg cggacagcgg ggcgtcggtg ctgctcacac agacccgcct
35221 gctggagcag gcggaagctt ggcgcagcga cggagctcta gtgctgcaaa cggtgcttgc
35281 gcttgacgac gccgcgacgt acagcctcgg agtggctgtg ggcgaagcag gcgcagaggc
35341 ggaggctttg gcgcaagcgg aaacggcttc tgccgaaacg tccgccacgg cagaagccga
35401 gaagaacgta ctagcggcgg atctcgcatc gaatccggcg aatgtgaaca agccgcgcga
35461 tttggcttat gtcatctaca cctccggtac gacgggccgt ccgaagggcg tggcggtgga
35521 aaaccgcagc ctggtgaaca cggcggcggg ctaccggcgg gactaccgcc tggatcagtt
35581 cccgatccgg ctgctgcagc tcgccagctt ctcgttcgac gtgttcgtcg gcgacattgc
35641 gcggacgctg tacaacggcg gcaccatggt catcgtgccg aaggacgacc ggattgatcc
35701 aacccgcctc tacggctgga ttcgcgacta cgctgtgacg gtgttcgaat cgactcctgc
35761 gctgatcgtg ccgttcatgg agcatgtgca tgccgagggt ctggacctca gctcgatgca
35821 gctgctgatc acaagctcgg atgcgtgcag cgtagcggat taccgcacct tgcaggagcg
35881 cttcggctcg cagttccgta ttattaacag ctacggcgta acggaagcgg cgattgactc
35941 cagcttctat gacgagccgc tggagaagct gccgaagacg ggcagcgtgc cgatcgggaa
36001 agcgtggctg aacgccaagt tctacatcgt ggatgcgaat ctgaagccgg tgccgatcgg
36061 ggtgttgggc gagctggtta tcggcggagc gggtgtggcc cgcggttact tgaaccgccc
36121 ggatttgacg gcggagaaat tcgtagacag cccgttcgcc gctggggagc ggctgtaccg
36181 gacgggcgac ctggcgcgct ggatgccgga cggcaacgtg gacttcatcg gccggatcga
36241 caaccaggtg aaaattcgcg gctatcggat cgagcttggt gaaattgaag cggccatgaa
36301 aaatttgcc ggtgtgcgtc aagcgcttgt catcgaccgg atggacgagc gggggcagaa
36361 atatttgtgc gggtatgtcg tagcggattc cagcttcgat ctggaagggc ttgtggccca
36421 tctggacgcc gcgctgcctt cccatatggt gccttcgcgc atcatgcgcc tggatcaaat
36481 gccgcttacg ccgaacggga agatcgaccg taaagcgctg cctgtaccgg aaggaagcat
36541 tcgtgccgag gctgcataca cggcgcctcg tactcctgcc gagcaagcac ttgcctcggt
```

```
36601 ctggcagtcg gtgctgggcg tggatcaggt cggcacgatg gacaatttct ttgcgctcgg
36661 cggcgattcg atcaaggcct tgcaggtatc gtcccgtctt ttgcaaacgg ggtacaagct
36721 ggtcatgaaa gatttgttcc attacccgac gatttccgcc cttagtttgc agctgcaaac
36781 ggcggagaga acggcaagcc aggccgaagt gacgggagag gtcatcttga ccccgattca
36841 gcgctggttc tttgaacaaa atccggccga cgtgcatcac agcaaccagg cgttcatgca
36901 gttctccaag caaggctttg acgaagaagc tttacgccaa gcggtgcgtc agcttgtcgt
36961 gcatcacgat gctctccgta cggtttaccg ccaaactgag aacggctata ccgcctggaa
37021 ccgcggcgcc ggggagaacg aagcgctgtt cgatctggaa gttgtcgatt tcaagggagt
37081 cggcgacgtg aaagaagcgg tagaggctaa ggcgaatgac attcaagcga gcatcgatct
37141 ggaaaacggt ccgctggtga agctcggctt gttccgctgc gacgacggcg accacctgct
37201 catcgcgatc catcacttgg tcgtagacgg cgtatcctgg cggattctgc ttgaggattt
37261 tgctgccggc tatgagcagg cgctgcaagg gcagccgatc cgtctgccgc tcaaaacgga
37321 ttcattccaa acgtgggcga aacagctcgc tgattatgcg aacggtccgg cgatggaaag
37381 cgagagagag tattggcagc atatcgagca attgatctat gagccgcttc cgaaagattt
37441 tgaacaaggc agatccaagc tgaaggacag cgggctcgtg accgttcgct ggacggcgga
37501 ggaaaccgaa cagctgctga agcaggcaca ccgtgcttac catacggaaa tgaacgattt
37561 gctgcttgcc gcgcttggcc tcgcggtaca agcttggagc ggccgggaac gcgtgctggt
37621 gaatctcgaa ggccacggcc gggaagatat tttgccgaat gtggacatta cgcgcacggt
37681 aggctggttt acaagccaat tccctgtcgt tctggagccg ggtcacgccc aggcgctcgg
37741 tcatcaggta aaacaggtta aagaaagctt gcgccgcatt ccgaacaaag gaatcggcta
37801 cggcatcctg cgttatttgt cggcgccgcg tgacggcgag cgcttcgctt tggagccgga
37861 gatcagcttt aactatttgg gtcagttcga ccaggattac gaaagcagcg gctcgaggcc
37921 gtctccgttc agcccgggct ccgattcaag cccggatgca gtgatggatt atgtcctcga
37981 tatcaacggt atggtgtcgg aaggagtgca ggaactcacg atccgttatg ggaaacccca
38041 gtataaacgg gaaacggtag agcgcctggg caccctgctt cattcgagct tgcgtgaagt
38101 catcagccat tgcgtatcga aagagcggcc ggagcttacg cctagcgacg tactgcttca
38161 agatgtgacg ttggaggaac tggagcggct ggccgaacat acggcggcgc tcggcgaact
38221 ggagaatgta tacaccctga ctccgctgca aaaagggatg ttgttccaca gcctgctgga
38281 tgccgattcg gaagcttact tcgaacaggt gaccttcgat ctgtacggaa gcctgaatgt
38341 cgaagccttc acccaaggat tggatacgct ggtgcagcgg aatgaggcgc tgcgtaccaa
38401 ctttattacc ggctggagag acgagccgat tcaagtggta ttccgcgagc ggaagtgtga
38461 agtgtacttc gaagatattc gctcggtaag cgatgaagac ccggagaaga cgatagccga
38521 tttcgtcagc gcggataaag cgaataagtt cgatttggcc cgaggctctc ttatgcgcgt
38581 gacagttttg cgcacgggcg acgagtccta ccatgtgatc tggagtcacc atcacatttt
38641 gatggacggc tggtgcatgt ccttcatgat caaggaagtg ttcgacacct acttcgcgtt
38701 ccaggagaag cggacgctgg agcttcctcc ggttacctcg tactcccggt atatcgaatg
38761 gctggaagct caagatgccg cgaaagcttc gcgttactgg tccgagtatt tggcgggtta
38821 cgatctagta gttgttgttg gctgctcgta tacaacaacg gactggctct caggctccca
38881 tttaacattt gtccccaagg cctctccgat tacccttaca ggtaccattg tcgtcccatc
38941 cactatggat gctggctgct cgatagaata cggctgaccg ttaatatatg ctttcggact
39001 gttaacaaat aatttgacgg acacgccgtt acgggtgaca gtaacagcct tttctttcgg
```

39061 atc

Appendix B: The nucleotide sequence of the insert DNA of cosmid Col-19

Sequence 36405 bp; 9885 A; 7760 C; 9140 G; 9620 T

```
   1 gatcgtagtt caatgtcttt tgaatttctt gaggagtagc tccttcttgc tccatttctt
  61 ttctggtatt aaattgtcct tgggaaatcc agttcacggc aggcgataca aggatgctga
 121 atgcaatatc atgttcctct ttaacgatct acaactacgc gcttcatacg catcatgccc
 181 tcattatttt cgataaaaaa ggccgttaat cccgtaataa tattaatctg agtgttgatc
 241 gtggagtcaa tttctgcatt ggcagcccga taggatgaaa attgtgtgag aatatccatc
 301 acttcctgcc gattttcaat caaataattg aggcatttca gaatagattc atgccttttc
 361 agaaacagtg tgttttctgc cagttctgga tttatactca tggcacattc tccctttcct
 421 gtatgcagtt ttggtatata tcttgttatg ctttgcgaga agcagctttc tctagagcaa
 481 cacgctccat ctccagacga tcactggaaa ggacggcaag tcggtggaac ttgaacaaca
 541 tgatggccaa tggagtatga gcaagccacg gacttaccca ttaaaccaat atacaataga
 601 caattggcta gcagcgattc aaaatgtaaa gttgggtgaa gtggtggaat cttctcctac
 661 agacgtagcc aaatatggga tctccattac aggaggtagt cgggcttcta aagacgatgt
 721 caaaaatgcg gcaagcaaaa aacaaaatcc tattcaacgt gctatcaaga cgctggcaga
 781 tatttttatt cccatcttgc ccgcaatcgt gaccgcaggt ttattacttg gaataaataa
 841 tattttgaca ggcaagggaa ttttcttcga aggtatggct ctggttcagg tctatcctca
 901 atggactgac tttgctgcta tcattaacac cattgccagt accgcttttg catttctacc
 961 tgccttaatt ggatggtcag ccgttgtccg gtttggcggc agtcctttac tcgggatcgt
1021 actgggtttg attctggttc atccggatct gctaagtgca tacggttatg cagatgctgt
1081 taccaaagga accgtcccgg tttggaattt gtttgggtgg gaagtgaaca agattggtta
1141 tcaaggacag gttttacccg ttttagtatc cgcttacgtg ctcgcgcaca ttgagaaatt
1201 cctaaataag aagatagcag attcatttaa actgttactt gtagcaccga ttacattatt
1261 aataacggga tttctggcat ttacagtaat cggtccggtt acttttatga ttggtaatgc
1321 tattacgtct ggtcttgtat acgttttcga tttggctcct gcgctgggag gcctgattta
1381 tggaggcttg tactccttgc tcgtgattac gggaatgcac catactttcc tggcggtgga
1441 cgtgcaatta attggtacgc agggtggaac attcctttgg ccgatgctgg cgctgtccaa
1501 tatcgctcag ggtgccggag cactggcaat gatgctggta ttgaaggagc aaaagggaaa
1561 aggattggct cttacctctt ccgtgtctgc atttctgggt gtaactgagc ctgccatatt
1621 cggggtgaat atccgatata aatatccgtt tgtattcgca atgatcggtt ccgctattgc
1681 aggtatactg ctcaccgtaa atcatgtgat ggcttcttcc attggtgtgg gcgggattcc
1741 aggcttcttg tctattttcc caaacaactg gggtgtattc ttcatcggga tggcgatcgt
1801 gttaattgtg ccatttgctt taactttaat gtatggcaag ttgttaacta agcgccgtca
1861 ggctgcacca gccaatgatt ccgtacggaa agattcgctt ccgcaaaata ccgcaataga
1921 tgtgcttgaa gtgaggactc ccgtacgagg acgtgttgta ccgctggaac aagttcctga
1981 cccggctttt gccgaacggc agatggggca aggagttgca gtggaaccga aggatggcaa
2041 gcttcttgcc ccattcagcg gaaaagttat tcatatcatg atgaaaagca agcatgcctt
2101 gattttggag cattcgacag gtgtacagat tctggtacat gtggggatca acacggtgtc
2161 actgaagggt gagggattca ccatgcatgt ggagaccaat caaaccgtgg agcaaggaca
2221 actattgatg gaatttgacc gtgatctgat tcagaaatcc ggatatcccg tcattacacc
```

```
2281 agtcattgtc ccggatggtc agagtatgat tgaatcggtt gaggaactgc caggtgaggc
2341 agacgatttg cagtatacct taatgagaat tcatttgagt tcttcatcag acaatacgaa
2401 ttaagtcaga ttcatcgtac agcctatttt agtgaacagg aagtgaaagg atgctacaat
2461 gaatgtggtg atgaaaattg acaaagaata tatttttgca aatttatcag gagtactctg
2521 gacgaatcca gacaggcgaa ttggctccag gcacgaaaat gccatctgaa aatgaactgg
2581 cccgagagta caacacttct cgcgaaaccg tgcgtaaggc acttcacttg cttgcacaaa
2641 atgggtatat ccataaaata aaagctaaag gatcttttgt tctggatatt ggacggatgg
2701 attttcctat aggaggattg gtcagtttca aggaaatgtc tgaacggatt gggcgtacat
2761 cacgaaccat tgtacatgaa aatcaactaa ttccagttcc agatgacata gctaatcatc
2821 ttgagattga caaagacaag tatttggtct ggaaagttat tcgtgcgcgt gagattgaag
2881 gtgaatgcgt catactggat aaagattact ttcgatcaga cattgtacct gtgcttaccg
2941 aagaaattag tcgtgggtcc atctacgaat atttggaggg agagctccat cttcaaatta
3001 gttatgccaa aaaagtcgtt tccgtagaag atgcagacga agaggatcat agactgatgg
3061 atcttcagaa ctacactcat gtagtcgtag tacgtaacta tgtttatctg gaggacacca
3121 cgttgttcca gtataccgag tcaaggcatc ggctggacaa attccaattt gtcgactttg
3181 ccagacgttc tcatggggag acaagagagc agattttttt aaataacgat caataaaaca
3241 gttaagaacc ccgtcttttg aagacgaggg ttctttttta ccttggatgc aactgctgag
3301 cagcttgtat actggcattt gagggatttg aatttgcgtt ttatccaaaa gctgtttata
3361 gaatattaaa aggttatata ctgtttagaa tacaattcca tgatcaatgc aatggaaatg
3421 aggtgtatgt tatttatgta tagagtggct atctgcgatg atgagcacaa gcacagagaa
3481 cttgttaaaa atatgctgat tactttgtcg ataaagacca atatagaatt taaagttgaa
3541 ttattcgact caggagaaca gctggtatct cattatcaaa accatgaatc tccattccat
3601 atcttaatat tggatgttga aatgaatggc ataaacggaa ttcaaaccgc ccaaaaaata
3661 aggaatttaa atcatctaga tgaacagatt gtttttttaa cgagctaccc tgaatatatg
3721 gtggaaagtt ttgacgtagt tacatttcag tatttaataa aaccgattgc tccccatatc
3781 tttgaggaaa aaatcataaa gctgtgtcaa tattttcaat ctctcgataa aaaattcgtt
3841 atcatcaagt cagcttatga ggaagtgctt ttaaaatatg atgatatcat tagtattgaa
3901 gtggccaaaa gcttaactat aaaaaacaaa ctacattttg taacctctac tcaaacctat
3961 gatagcaaag ggattctttc ggactatgct tcagcattaa aagaacataa ttttctacaa
4021 attcatcgtt cgattattat aaacttaatt catgtgaaaa aatttgctgg cggtgatgtt
4081 ttgatgtcaa atgatgtgaa gttacctatt gggcgttcta aaataaaaga agttaaagat
4141 cgtaaagcaa acccaacaag tggattatcg catcatggag attacaggaa ttaaccctgc
4201 aacctttggt ggcgcttcga tggacaagaa cgggaacgat tcgaaatact attacatgta
4261 caataaaccg atcaaacgag tcttggattc ggtgcggctg ctgctctggg agcaagcgtg
4321 gcaacacaac gtccagtagt ggcacttgta ggtgacggtg cgttcaatct cttttgttcg
4381 gattttatga cggcggtgga gtatcaaata ccgttaatct atctcgtact taataatggt
4441 gggtatggtt ggttacagaa tcagatgaat tataagcatc tcacagggca gtccctcccg
4501 tttgtttctg aggcggccaa gacaggaatt caaattcctc agtacgagtt tgtggaaagt
4561 cttgcaattc ggagtaaggg agaggcacgg gaacagcttc agcttgcttg gcaaaagtat
4621 caggagaaca agcttgtgat cgtcgaggta tatgctgatc ttgcggatgt gcatgaaaaa
4681 atcagtcatg tgtatggcga ctttcctctc tacgagcaac aggcttctaa gaattaattt
```

```
4741 caacgaagga cgtggataca gtgagaagac gtgttgtagt gactggacag ggagtcgtaa
4801 cgccgatcgg gcttaatctt acggatttct ggaattcact gttggaaggg aaaagtggaa
4861 tcggacctgt cactgctttc gatacagcgg acattccgac aaagatcggt gcacaggtta
4921 cgggctttga cccattgcaa tatatgagca agaaggaagc taataaagtc gcacagttca
4981 ctcagtttgc tttggcggct gcccggcagg cagtggagca atcaaaatta gtgattgatg
5041 agcggaatgc aagtcgcgtg ggagttattg tcggctctgg agcaggcggc ctcgacgtta
5101 tagaggagaa ctttgggaaa ttaacaaagc ttggtcccaa aagagtttct ccttattttg
5161 tttcctcgat gatgattaat tcagcacctg gtgaaatctc cattgctatt ggagccaagg
5221 ggccttcttt tgctgtggtt acggcttgtg ccacaggaag taatgccatt ggtgaggcat
5281 tacgtacgat ccagtacggc acggctgatg tcgtggttgc cggtggagca gaagccaatt
5341 ttagtcaact tgatctggca tcttttgcga atatccatgc tctctccaga cgtaatgaag
5401 agccgcagaa ggctagtcgt ccgtttgatg tggatcgcga tggattcgtc atcggtggag
5461 gagcaggagt cttggtgctg gaggagttgg agtatgcttt gaaccgcgga gccactattc
5521 tggccgaagt tgtgggttac ggatgcacaa ctgatgcgta tcatatcacg gctccagatc
5581 cctcaggctc cggcccggca gaagcgatca atatcgcact gcaggatggc ggcttaacac
5641 cgcaggaaat cgactacgtc aatgcccacg gtacgagtac gccatttaat gatcgtatgg
5701 agattaatgc tattcaaaaa gtattcgggg agcatgcccc ccgtttggcg atatcgtcta
5761 ttaaatccat gactggtcat ttaatgggcg gtgcaggggc agtcgaatta attgcgactg
5821 tgcagagtat gattcacagc aaagtaccac cgacactcaa ttgcgacaat cctgaggctc
5881 ctgaattgaa ctttgtgccg catgtatatc aggaacgcga ggtccgcgcg gcgattagca
5941 attcatttgg tttcggaggg cacaacgtct gccttgctgt tcgcaagtgg gaaggggaat
6001 aagcctgtga gcgaagagct aatccatatc ccggatgtgc ttccgcaccg gtacccattt
6061 cttctgttgg atggcgtgac cgggtacgaa acgtcgcaat gggcgaaggg atacaagctt
6121 gtcacctgga gtgagtggtt cattactgac actaacccat acatgcctcc tatgttgatt
6181 gtggagtcgc ttgcacaatt agcggcattc gctgcgattg ataaaaaagg catttcatac
6241 ctgactaacc tcaatggaat tcgattccac agcttggcca agcctggaga ccgaattgat
6301 ctctattttg aagtgatcaa gcgtcggcgg ggctatattc tcggaagcgg tcgggcttcc
6361 gtaggggatc gattggtggt tgaggcggaa gaaattgttt cactggatca atgacggggt
6421 ggaagcgtac gttaaactgt ccttttgtcc tcctgactgc caggtgcagc gggaggatat
6481 tcacatcatg gaataggagt gaatgagatg aatgacatgc agttatatga tttaacaaat
6541 gcgcagaagc gtatatggta taccgaatta ctctacccag atacgtcagt gtcacagctt
6601 tccggtacag ctaagatgaa ggggcatatc aatattgctg cctttatgca gtccattaat
6661 ttgattatca aacagtatga tgcgttccgc atccgtatta cctcagtgga tggagtgcct
6721 cagcagtacg tcgttcctta tgaagagaga cagttggagt gcctggatct tagtcactat
6781 gaaagtgtat ctgaggtgga agccttactt gagcaacaca aaagaaaacc cttgcccctg
6841 ctggattctg agctcttcca gtttttaatt gtgaagatta gcgaggaaga gtattggatt
6901 aatatcaaga tgcaccatat tatttctgac gggatctcaa tggtggtcta tggcaatcag
6961 ctgacagcat tttacatgga gttaattcaa ggaaatgaac cgaagctggg cgacgattgc
7021 tcgtatattc aatatattgc agatgagaat gcatacgaac tttctgacag ataccaaaag
7081 gataaggctt actggctaga taaattttcc gatttgcctg agcttacggg ttggaagtca
7141 tataatccgt tatctttaag cacccacgcc gttcgggagc attttaccgt accagaagtg
```

```
7201 ctatatcacg agctgcaagc attttgccaa cagaatagga tttctttgtt ccagttcttc
7261 atgggtgcga tgtatatcta catacacaaa atgacgaatc agccggatgt ggtgattggc
7321 acttcgttcg ctaaccgggg gaacaaaaaa gagaagcaaa agataggtat gttcgtcagc
7381 accgctgctg ccagaacata cgtcaaaaag gatatagatg tgttgagctt cctgcaggat
7441 gtagccagag atcagatgtc agtcctgcgg catcagaaat atccatacaa tcagttaatt
7501 caggatctta gagaaatgca tgggaacaag gatattcagc ggctttttgg cgtttcaatg
7561 gaatatcgtc ttatcaattg ggttgatttg gatgatgtgc gtattttgac agattatgat
7621 ttctgcgggg acgaagtgaa cgatttcgtg tttcatatcg tggagatcct ggatgaaggc
7681 gaactggtac tggatgtcga ttatcggaca gagctgtttg aacgcagtga agttaaggac
7741 atggtttccc agttgcttac gatcgccgag cagatcattc attcacctca gctttctatc
7801 gcagaggtaa gcttattagg tgaaccagaa gagcaatcca ttttggctct ttcggaaggc
7861 gccgcagtcg attatccacg agagaagacc attcatggct tattcgagga acaagccgag
7921 cgcacgccag atcacgtagc cgttcagatg gacgagcaga gcattacata cctagctcta
7981 aacgagcagg ctaaccagct tgcgagatat ttgcgctccg agggagtagg ggcagatacg
8041 ctcgtaggga ttatggctga ccgttccttg gagatggtca tcgggatgtt ggccattttg
8101 aaagcaggtg gtgcctatgt accaattgac cccgattatc ccgaagagcg tatccactat
8161 atgctggagg attcaggtgt aagtctgttg ctcacccaaa gtcatctatg ggagagcacc
8221 acttttgacg gaaagcttgt gagtctggac gaagctgcaa cgtatacagg agatgcttcc
8281 aatctggaga gtatttcggg accaagccat ctggcctatg ttatctacac gtcgggtacg
8341 actggcaagc cgaagggcac gctgattgaa cacaaaaatg tagttcgact gctctttaac
8401 gataaaaatt tatttgactt caactctcag gatacgtgga cgctgttcca ttcgttctgc
8461 ttcgatttct ccgtttggga gatgtacgga gcgcttcttt acggagggaa attggtgatt
8521 gttccatctc tcacagccaa gagcccagca gctttcctgg agttgttgaa agacaaccaa
8581 gtcaccatct taaatcagac gccgacgtat ttttatcagg tgctacagga agagttaatg
8641 cactcttcga cagagcttgg ccttagaaaa atcattttgg gtggagaggc tttaagtcca
8701 tctcttctga ggaactggcg ggtcaagtat cctgatgtgc agctgattaa tatgtacgga
8761 attacggaaa caacagttca tgtcacctat aaggaaatca cggaacatga gattgaagcg
8821 gggaaaagca atattggcag aacgatcccc acacttagcg cttacattct cgatgagcaa
8881 agacgcctgc agcctgttgg ggttccagga gagctataca ttgcgggaga cggtcttgcc
8941 cgtgggtatt tgaatcggcc ggatttgacg tctgagaaat tcgttgagca tccgtatcgg
9001 gtgggagagc ggctgtaccg aactggggat cttgctcgtt ggttgcctga tggcaatatt
9061 gaatatttgg ggcggatcga ccatcaggtc aaaattcgcg gctaccgaat tgagcttggc
9121 gaggtagaag cccaaattct caaggctcca agcgtacgag aaacgattgt tctcgcacgg
9181 gaagacgaac agggccaaaa attgctgtgc gcctactatg tagcctccag tgatctttcg
9241 ccggggggaat tgcgggctca gctggcggcg gaactccccg cttacatgat tccttcttat
9301 tttgtcccgc tggagcaaat gccgcttaca ccaaatggca aactggatcg ccgtgcgttg
9361 ccggctcctg aaagcagcgt acaatccggt gaggcttatt tggctccgag aactgctgtg
9421 gaagctcaga tggtactcat ctggcaagat atcttgggcg ttgctcgtgt cggtgtcaga
9481 gataatttct ttgaaattgg tggtcactct ttgcgggcaa cagtgctcgt ttcacggatt
9541 cacaaagaat tgggatgtag catttcgctg cgtgaggtgt ttcagtcacc tacggtcgaa
9601 tccttggcgc aacttgtgaa aaaacacatt ccgaccctgt acgaatccat cccacaggca
```

```
 9661 acggaaagcg aagcttaccc agtgtcctca gcgcaaaagc ggttatacgt gctgagacag
 9721 atggacgggg gagagctcag ctacaatatg ccaggggtct tcacagtgga tggactgttg
 9781 gatcgcacgc ggctggagtc tgcgttccag gcactgatcc agcgtcatga atccctgaga
 9841 accggctttt atatgcagga tggagagctt gttcagcgtg tgcataggaa tgtgtcgttt
 9901 gctttgaact acacagaggc tttggtggag gagacggata cgctcattca caactttatt
 9961 cgtgcctttg atctgagcca ggctccatta ctgcgtgtta gcttggtgaa gctccaggag
10021 gagcgtcatc tgttgctgtt tgatatgcat cacatcattt cagatggggt ttctattcaa
10081 attttgatag aggaacttac tcatttgtat caaggagaac agctaccaga gctgcacatc
10141 cagtacaagg attatgccgt atggcaacgg gaacagtcag agaaccaatg gcaagatctt
10201 gagaaatatt ggttgcaatc ctttgaagga gagttgccgg tattggattt gcctacagac
10261 ttccaacgac cttcagttcg gagcttcgag ggtagccgaa ttgattttac attggatgag
10321 tctggcaata aggcgataca agagcttgca tcccgtacag gtactacact gtatatggta
10381 ttgctggccg cttattcggt actactgcac aaatatacag gacaggagga catcgtcgta
10441 ggttctccag tagccggaag accgcaggct gagcttgagg gcatcatcgg aatgtttgtc
10501 aacacactgg ccttgcgtag ctacccggca ggagataaaa cctttcagga ttatcttctg
10561 gaaatcaagg aaacggcgct caaggcgttt gagcatcagg attacccttt tgaaaaattg
10621 gtcgaaaagc tgggcgtagg acgtgatgtc agccgcaatc cgctctttga caccttattg
10681 gtattacaaa ataccgagca ggaagagcag gatatggacg gagtgcactt tactccttac
10741 ttgatggaca ccgtcacagc caaatttgac ctatccctca atgtagagga gaagggatca
10801 aaactagcct ttggccttga gtatagtacg gctttatatc ggcgtgaaag tgtagagcga
10861 cttgcaacgc acttgctccg gattctgcac gcagtcgcgg ccaatcctca gttgcaattg
10921 gccgagattg aaatgatcac accggaggag aaagtacaga tcgttgaagt atttaacgcg
10981 acatcggctc cttatccaag ggacaagacc attcatgagc tgttcgcaga acaagtcaag
11041 cgtacaccag agcagacggc gcttgtattc ggcgatgtcc agctaacgta tcttgaattg
11101 gaagacaagg cgagccgact ggcccaaaca ctgcgtcgtt tgggaacgtt gagggagcag
11161 cctgtggccg tgatgggcgg acgaagcatc gagatggtca ttggtatgct cgcggtgctt
11221 caagcgggtg gagcctatgt gccgattgat cctgattacc cggaagatcg ggttcgttat
11281 atgcttaatg attccgacgc caagctatta ttggtgcaaa agggcgagct tataagtgtc
11341 gactacggta taccgattgt cgatcttagc agtgaagagg cctatgcagc tgagcctgcc
11401 cagccggaga ctgctcaggg atcgcagggg cttgcttatg tcatctatac atcgggtacg
11461 acgggtagac cgaagggcgt tatggttgaa caccggaacg tggtccgtct agtcaaagag
11521 accaactatg tggagctgaa tgaatccaca cgaattttgc aaacaggagc cgtagccttt
11581 gatgcttcta cattcgaaat atggggagcg ttgctcaacg gtgggcagct ctatttcgtg
11641 gagaacgacg atatcctgat tgccgatagg ctcaaagcgg ccattaccaa gtacggatt
11701 acaacaatgt ggctcacttc accgcttttc aatcagcttt ccctgcagga tgagtacctg
11761 ttcagagggc taaaagcatt gctagtcggc ggtgacgtac tgtctctatc tcatatgaat
11821 cgtgtaatcg aggctaatcc tgatcttgtc cctatcaatt gctatggtcc gacagagaat
11881 acgaccttct ccaccactta caagattcca ggtcgtgccg aagggggtgt gccgattggt
11941 cgcccaattc gtaattcgac cgcttatgtg gtcaatggat cgctgcaatt acagcctatt
12001 ggtgcttggg gtgaactcat tgtcggcggt gaaggtgtag cgcgcggata tctcaatcgt
12061 cctgatctca cagcagagaa atttgttcct agtcctgtga aggacggaga accctgctac
```

12121 agaactgggg atttggtacg ctggcttcca gatgggaatt tggagtttaa aggaagaatt
12181 gatgagcagg tcaaaatacg tggttaccgc atcgaactcc ctgaaatcga ggcccaactg
12241 gccaaggtgg agtcagtaat cgacgccgta gtggtcgttc gcgcggatga gcttggcgag
12301 aagcagcttt gcgcttatta tgtggcggat cgtacgctca cggcaggcga agtacgtctt
12361 tccctatcgc aggtacttcc aggctatatg attccatcct actttatcca gatggatcgt
12421 atgccattaa cgtcaaacgg aaaagtggac cgcaggtctc tgccggctcc tcaagtaggc
12481 gcgcatacag gacggaagta tacagctcct cgtacaccgg ctgaggaagc tttggcatct
12541 gtctggcaag ggtgctggg tgccaaacag gtgggtatcc atgacaattt ctttgaattg
12601 ggtggagact ccataaaagc tattcaggtg tcgtcacggt tactgcaggc tggctatcgg
12661 ttagagatga agcagctgtt caaatcgcca accattgccg agctaggcgc ggaaatacaa
12721 acggctgtgc atatggctga acagggagtt gtgcgtggaa cgactcgctt gactccagtc
12781 caacagtggt tctttggacg gaagcaggca gaacctcatc acttcaatca agcggttatg
12841 ctgtatcgtg aacagggatt tgaagaaaag gccttgcatc aggtgctaag aaaactcgct
12901 gagcatcatg acgcccttcg catggttttc cgtcagacag agcatggcta cgaagcttgg
12961 aatcgtgatc ttgatgaagg agagctttat agcctgttca ccgctgattt acggaatgaa
13021 tccgatccga ctgcagccat tacatcgctg tcggatgaca ttcagcgcag tatcaatctg
13081 gcagaaggcc cgctgctgaa gttaggactt ttccattgtc aggatggaga ccaccttcta
13141 atcgtgatcc accatttggt ggtggacgga gtatcctggc ggattttgtt cgaggatatt
13201 gcagcaggct atgagcaggt gattcaagga caagcgctga cattcccgca gaagacggat
13261 tccttccgtg actggggaga cgcccttgct cgttattcgg aaggtcctga tatggagact
13321 catcgggcgt attggagaga gctggaggat cagccactcg aacagttgcc gaaggatgag
13381 gctgtggaaa gccttctttt acaggatagc aaagtagtaa cagcacaatg gactctagaa
13441 gaaaccgacc aattgttgag aaaagcccat cgtgcttatc aaacagagac gaatgatctg
13501 ctattaactg ctctgggcat ggcggtatcc aaatggtctg gcatcggaaa ggttgctgtg
13561 aatctggaag gacacggtcg tgagccgatt ataccgaata tcgacatcac ccgtaccgta
13621 ggctggttta caagtcaatt tccggtgatt ttagacttgg gggatgaccc ggaagtggcc
13681 tccttgatca agtctgtgaa agaggggcta cgccggattc cgaacaaagg tattggctac
13741 gggttgctca aaacgatggc aagtcagttg gatgaagaca gcttcagctt acagcctgag
13801 atttctttta actatctggg gcaatttgat caggatttgc aaggaagctc gttgcagatt
13861 tctccttatc cgaccggaag cgcccaaagc ttgttggagg aaccggccta tacgctagac
13921 atcaatggca tggtgacgga cggagccttg actctgacga ttacctataa cggaaaacag
13981 tataagtcat ctacgatgga acagctcgct ggatatattg aagaaagcct gcgggagctt
14041 ctccagcatt gcgtaaacca agaaaaaacc atattgacac caagcgacgt gcttgcgaag
14101 ggtctaagca ttgccgatct ggaggagctt tctaagcaga cgagtcacat tggcgatatt
14161 gagaatgtat atagtctgac accgatgcag aagggcatgc tgttccatga tatgtttgag
14221 ccgcatacag gtgcttattt tgagcaggct gcctttgact ttaagggtag ctttgatccg
14281 accgccttcg gacacagtct ggatgcagtg gtggagcgtc atgctatcct gcgcacgaac
14341 ttttacagcg gatggggcag cgagcctttg caggttgtat ttcggcacag aggcgctaaa
14401 ttggtgtacg aagacctgcg ggagatgaat gcatcgcagc gcgaagctta tctgaagaca
14461 tttggtgcta aggacaaagc actgggcttc aacctagctg aagacgagct tctccgtgta
14521 tcaattctac aaacagatga agagagcttc cgtctcttat ggagctttca ccacatcgtc

```
14581 atggatgggt ggtgtgttcc gttaattacg caggaggtat ttgaacacta ttttgccctc
14641 ctggaaggaa gagagccgca gttggcagag gttcatccgt acagtcgata tatcgaatgg
14701 ctggaacagc aggatgaagc agctgcgtcc aactattgga gccgatatct ggccggttac
14761 gagcagcaga cgcttttacc tcaagtcggt ggagcaagta agggagaagg ctatgtagca
14821 gaaaagctga attatcctct cagcagggaa ttgactgagc gccttgaaaa ggtggccagg
14881 gatgcccatg tcacgatgaa tatattgctg cagtccctct ggggcattgc gcttcaacgc
14941 tataacggta gcaaggatgt cgtgtacgga agtgtagtat caggcagacc agcagaaatt
15001 ccgggcattg atcggatgat cggtttgttc atcaatacga ttcccgttcg tgtgaagaca
15061 gaggagaatc tcccctcac agttctgatg aagcagcagc aggaacaata tatggcttct
15121 catatgtatg acacctaccc gctgtttgag attcaggctc agacggatca gaagcaggat
15181 ctaatctccc atattatggt gtttgagaac tatcctgtgg aagaggaggt agagcgtctg
15241 ggtggtggcg aggctgcctt tgagattgag gaagcggagc ttcttgagca aacgaattat
15301 gattttaatt taattgtcct ccctggcgaa gaaatgagat tgctgttcca gtacaatgca
15361 cttgtttatg acccagtgac aattgaacaa atcaagggcc atctggttca cctcatggaa
15421 caaattgtag agaaccctgc catttccgtg gatgcactag aattagtcac gccgcaggag
15481 agagaacaga ttctgaacgt atggggaaat acaaaaggca tttacgagca ctgtaacacg
15541 ttccacgggc tgttggagga acaggcggga cgaacgccgg atgcgactgc catttggttc
15601 gaggacgaga gtctgaccta tgccgagctc aatgcaaaag ccaatggact ggcgagaagg
15661 ctccgtactc agggaatcaa gacgggagat ctggtgggac tgattgctga acggtcgctc
15721 gaaatgatcg ttgggatcta cggcattatg aaagccgggg gtgcctatgt tccaatcgat
15781 ccagagtatc cgaaagaacg aatcagttac atgcttgaag attccggggc gaagctgatc
15841 cttacacagg cccatctctt ggagcatctt ggatggacgg aaaatgtttt gctgctggat
15901 gaatcatcga cctatgatgc cgatacctcg aatttggagg atactgctgg cccggatgat
15961 ctggcttacg tgatctacac ttcaggtacg acgggtcagc ctaagggcgt attagtcgag
16021 catcggggac taccgaatct ttcggacgta tacgggacac acttcgaagt tacaccgcag
16081 gatcggatcg ttcagtttgc aagcctgtcg tttgatgcat cggtttcgga attttaacg
16141 gcgctgagcc acgggggtgt tctgtgcatc ccttctacac aagatatttt agatcttcat
16201 aagttcctcc tacgggtttt aatccttcaa gcacgcgttg acgcgggcct tatttaggtg
16261 atgtatatct ttaaaatata ttagtaatag taataggcgc actgaatatg tggataagtc
16321 actgtaaccc taataaaatc aagcctatcc acatgtgtat aaattgtgta taggctctga
16381 acttgttcag gttggatttt tgatttttc tattaggttg ttgataactt taaacagatc
16441 ctgatccact tgaatggatt tggagatttt ttcgtgcgca tgaatgacag tcgtatgatc
16501 acgcccaccg aaagcttcgc cgatttttgg caaagaatag tcggttagct cacgagacag
16561 gtacatggca atctgtcgtg gaaaagctac agccttcgtc cgcttacgcg ctttaaaatc
16621 ctccagccgt agattgtaaa attccccgac ttgatgctgt atatcctgaa tcgtgatcaa
16681 gcgatgtttt ctatggctaa ggccatttat gttcatcgaa accagatcat tccattgtat
16741 attcgtttat ccgtgactga cctgttccct ttaattcaca cagatgctga gcccaattta
16801 gatatggtgc tgtttatcgg cacggtacat tatctggttt cgggagaaga tgatgtatct
16861 aaagagccga tgaagctggg ggaagtagac aagttgcttt ataaaacgga ctctattcta
16921 gagtctctgt cgcagcggat cgtagctgac ccgtctaagc ggtatccatc tatttataaa
16981 gaaaattcgt tcaatgaatc ggtagatctt ttatcaataa tacccggttc tttttgagcc
```

```
17041 aatgatataa tgatgaaaag gaagatgttg gatgcgggat aatgcttcca catcagaaag
17101 ggtgttattt catatgccta catacaacaa actggtgcgg gatcggattc cgcacctcat
17161 tatgtctcaa ggcaaggaat gtcggacacg cattttggac ccagaggaat acaagcagga
17221 actgcgaacc aagcttcgtg aagaagccga ggaatatttt gaagccgaaa aagatcagga
17281 tgcgctggaa gaactggcgg atatgctgga agttatccga gcgttggcag aagaacatgg
17341 agcgaatgca gcagagctgg acaagctgcg ggcagacaaa gccgaggcac gcggcggatt
17401 ccaggagcgg gtgtttttaa ttgatgtcga tgaaggtaaa agttaactaa aatatttcaa
17461 agataagcac attcaaaagg ggagcttgat aactggagct accttttttt gtctcttttt
17521 aagcacccag tatggcgaaa attataaagt ggaagttatt ttaatttatt tacaagccat
17581 atgactattt agtcaacgag tttcatgtta tctcctccaa cttctacctc cctacacaaa
17641 ggttaaaatt ttataaaaaa agttcaaaac cagcaaagac ctttccgggc ctggaatgat
17701 accatgtgtt tgcacataaa aaaaggagag gcgctcattg cctcacttat gtaagcgcta
17761 tctaatttca actctaatca gcacttacct tctaattgac ccgcatcaca tgaggaaagg
17821 ggaacttcat atgaaaaaag gagctgttgt cgtatggctt ctggttttgg ccctgatgct
17881 ttcggcctgt aatttggcgg agagtggaac gggcagcaag gaaaaaggct atgtaggcat
17941 ttccatgccg accaaatcgt ccgagcgttg ggtaggggac ggggagaata tggttcggtt
18001 gtttcaggag caaggctata aaacggattt acagtacgcc gaggacgtgg tggaaaacca
18061 aatctcccaa atcgaaaata tgatcaccaa gggtgtcaat gttatggtga tcgcttcgat
18121 cgatggaaac acgctgacgg atgtgattca gaaggcacac gaccgaggca ttcaggtcat
18181 ttcttacgac cgtctgattc gcaacacacc gtatctgacg tattatgcga ccttcgataa
18241 ctttaaggtg ggcgtgctac aggcggcata tattgagaaa aagttgggac tcaaggaagg
18301 caaaggccca ttcaatatcg agctgttcgg tggttcaccg gatgataata acgcttattt
18361 cttctttaat ggtgcgatgt ctatacttaa accgtacgtg gattccggca aactggtcgt
18421 gcgcagcaag cagatgacga tggcgcagat cgctacactg cggtgggatg gagcattggc
18481 ccagtcgcga atggataacc tgctgagtgc gtactattcg gggggctaatc tggatgcggt
18541 actatcccct tatgacggca ttagcatcgg tatcatttca tccctaaaag gggtcggata
18601 tggatcagcg aacaagccgc tgccgatcat tacagggcag gatgcggagc tggcttccat
18661 caagtcgatt gtggctgggg agcagacgca aactgtattt aaggatactc gcaagctggc
18721 tgagcaaacg gtcgtgatgg ccaacagcat tttacagggc aagcaagcgg aagtgaatga
18781 caccgaatcc tataacaacg gtaaaaaggt cattcctgct tacttacttg atccgatctc
18841 ggtggatcgt acgaatgtgg aaaaggatat tgtcggcagt cactattaca ccaaagagca
18901 aatcggactg aaataagagt aagaatccaa ccgaaaggag cgtatcccat gactggaatc
18961 attttggaaa tgaagggaat taccaaaacc tttcccggcg tgaaggcatt ggaaaatgta
19021 aatctcaagg tcaaggaagg cgaaattcat tccctgtgcg gtgagaacgg tgcaggcaaa
19081 tcgacgctga tgaaggtact aagcggtgta tatccgcatg gaacctatga aggggatatt
19141 gttttccagg gcaaaacgtg tgagttcaag gatatcaagc agagcgagga attgggcatt
19201 gtcattatcc atcaggagct ggcgctaatc ccgtatctct cgattgcgga aaatatttat
19261 ctgggcaacg aacgcgccag tggcggcgtg attgattgga aagagacttt tgtaggaaca
19321 cgcgagctgc tgtccaaggt cggtttgagt gaaaatccca atacattagt tacgaacatc
19381 ggggtcggga agcaacaact ggtcgaaatt gcgaaagcgc tctctaaaaa ggtgaagctg
19441 cttattctcg atgaaccgac agcggcgttg aatgaggatg atagtgaaaa tctgcttaat
```

19501 ttgatgctgg aatttaaaag acagggcatt tcctgcattc tcatctccca caagctgaat
19561 gaagtcgcaa aggtgtcgga ttcaattacg attttgcggg acgggaagac gattgagaca
19621 ctggatatga aaaaggatga ggtcacggag gatcggatca ttagcgggat ggttggacgt
19681 gatctgacca gccgttatcc agagcgtcat gccgagattg gcaaggtcgt gttggaagtg
19741 aaggattgga cggtgtatca cgagcaccat gcagaccgca aggtgctgga tcaggtgaat
19801 ttgaacattc gtcgcggtga aattgtcggt atagccgggc tgatgggagc gggacggacc
19861 gagctggcga tgagcatttt tggcaagtcg tatgggcgca acatttccgg tcagctgatt
19921 aaagacggca agccgatcca gaacaacacg gttacggacg cgatcaacaa cggctttgcc
19981 tatgtgacag aggatcgcaa ggagtatggc ctcatcctga tggatgatat caagcgtaat
20041 atttcgttga ccggcctgaa caagctgacg aagggcgctg tcgtcaatga gcaggaggaa
20101 gtcgttgttg cagaggaaat gaaaaaaagc atgaacatca aagcgccaag cattttacag
20161 aaaacgggta atctgagtgg cggcaatcag cagaaggtag tgctgagcaa atggattttc
20221 gccggaccgg atattctgat tctggatgaa ccgacacgcg gcattgatgt gggcgccaaa
20281 tatgaaattt acacgattat tcaccgactt gccgctgaag gcaagggagt gctggtcatt
20341 tcgtccgagc tgcctgaggt actaggcttg tgtgaccgga tttatgtcat gaatgctggg
20401 cggatcactg gagaggttag ccgtgaggac gcgtcgcagg aaacgttgat gagatatatg
20461 accaagtcag gaggcgtgaa gcatggaaac cataactaaa ctatttaaaa ataacatccg
20521 ccaatatggc atgatgattg cgcttgtcgt cattatgatt ttgttcgagg tgctgacggg
20581 cggcctgctg ttgaagccga ttaatattac aaatctaata ctgcaaaaca gctatattct
20641 cgtgctggcc atcgggatgg tgctggtcat catcactgga catattgacc tgtccgtcgg
20701 ctccatagcc gcttttgtcg gtgcggttgc cgccataatg atggtcgatt ggcagcttcc
20761 agcatggatc gctgtcattg catctttgct ggtcggggcg ttaatcggcg cttggcaagg
20821 gttctgggtc gcttatgtac ggattccggc atttatcgtt acattggcgg gaatgctgct
20881 ctttcgcgga ttgacaatga tcgtgctgga aggccaatct atctctccat ttccgggggg
20941 attccaaaaa atcagttcag gcttcctacc ggatttttca tctaccggtt ataacctggt
21001 atccatcatt gtcggtattg cactcacggt atggtttatc gtgaacgaac ttcgcgaacg
21061 ccgttctcag cttaaatatg gatttgaggt tccgtcacaa acgttatttg tactcaagtt
21121 gattgtggtg gctgcggtta tcaatttgtt caccttcatg ctcgcgagct acgcggggat
21181 acctaacatt ctcgtgctgc tcttcatact gatcgttgta tactccttcg tcatgaaccg
21241 cacgattatg ggccgtcatg tgtatgcgtt gggtggaaat gaaaaggcag cgggtctgtc
21301 cggtgtcaaa acgaaaaaag taacgttctg ggtattcgtc aatatgggcg tgctggccgc
21361 catttccggt ctgatctttg ctgcacgtct gaacgcggct acgccgagag caggtacgaa
21421 ctttgagctg gatgctatcg cggcttgctt tatcggtgga gcttcggcgt ccggcgggat
21481 cggaaccgtt tttggagcca ttatcggcgg tttggtcatg ggcgttctga ataacgggat
21541 gtctctgatt ggtctgggta ttgactggca gcaaggcatc aagggcttag tacttctgct
21601 ggcggttgcc ttcgatattt acaacaaaaa caaaagatca gcttaagctc agattcgata
21661 tcgtaaaaca agacacctcc agtcccactt aaaatgtggg tgccggaggt gtcttgttta
21721 ttaggtcttg tcagctaacg caactaatct gtcaccattt gtccgccatt gacatgcagc
21781 gtttgccctg taatgtagga accatcatcc gaagccagca gcacataggc gggggccaat
21841 tcataaggct gcccggcacg cttcatcggc gtattggtgc caaaggtacg gatcgcttcg
21901 ggcggctgtg tggcggcatt cagtggagtc cagatcggac cgggagcaat ggcgttcacg

```
21961 cggatgcact gatcgaccag attattagcc agagcgcggg tgaagctgac aatggctcct
22021 ttggtagctg aatagtcgat cagattcttt tgaccaacat aggctgtaac cgatgccgta
22081 ttgacaatac tggccccagc acgcatatat ggcagtgctt ctgtggtcat atgaaagaag
22141 gaaattatat tcgtgtcaaa cgtgtcacgt agctgctgtt cgctgatatc tagatacgat
22201 tcgcgtacaa actgaacacc catgttatta atgaggatat caatacgacc gaaggtgtcc
22261 atcgtttttt gaacaaccag acagcagttg ctttttagcc ttagatcgct gggaattaac
22321 aaacaacgtc tgccaagcct gtgaatgaca tcacgggttt ccatagcatc tcgatgctca
22381 tcgagatagg caatgaccac atcggcccct tctttggcaa aagccacagc aaccgcacga
22441 ccaataccgc tgtctccgcc gctaataata gccacccgat ccagaagctt tccggtgccg
22501 atatactttg gattatcata aatagggcgt ggaaccatca ggtattccaa ccccggctgt
22561 ctatactgat gctggggagg aaaagctatc ggaacttctt tacattgggt ttctgtacca
22621 taataaggat attgccaatt cattgataat cccttcttta ggcgttttga gatccaatat
22681 attcaaaatg ggcgcaagcg gtgactcggg attatcgcct tatcccgaaa tcgtcgctcc
22741 atctgatgga atatggacac gtccctccaa accttcacgg ttgatatgtt ccgtaagctc
22801 tgagcgagtt agtagactat ggttgattgc ttccatgtgg acagccacga tttgtgtgta
22861 agaggcatgc tgcgcaacat ccgtcacatc ttgggcagtc ataatgatgg gatcaccaat
22921 cgtaaaccgt gctcctccag cgttcacaat tgtccactca ggacggtatt tatctagtgc
22981 ttccgcgact tcactgcacc agatcgtatc cccggccaaa tagacgacag gctcatctgg
23041 agcctgaaac acaaagccgg atacgggtcc catctgttcc ccgatttctc cggtgccgtg
23101 atgcccggta gtacggatga tctcaatttc tcccacggag ccagccaccg aaatcggagt
23161 aacctgtgta aaaccgtcag cagtaatttg ttcttcattt ccgggctggc agaagacggg
23221 aatttgtttg tccaaccagc gtactgctgc cgggtcccaa tgatcgggat gcaaatgtgt
23281 gacgatcaca acatcaggct gtccgagtgc ctgccattgc ggcggcaacg gaaccagcgg
23341 attacggcgg gcattttccg tattgggaac cggcggattt gcgccagcat cactcaacat
23401 cggatcaatc agaaatttta ctccagcgta ttcgatccat aaggttgcgt tgcggattag
23461 tgtgattttc atctatggaa gctcctttgg tagtgaatgt ttttggaccg agtgtatttt
23521 cggtggttct ttgtttctgt tacaatcata tctgacgcaa ggagctacag gctacggcct
23581 gctgaaagga aaaccagcat tatcgtttca tcatgaaagg aaaaataact atggaactat
23641 tcgtacctga tgacaccgat ctgcgcattc ttcatcattt aattgaagac agctctttat
23701 cgcacaagga gatcggtcag cttgttcact tgacgggcca agcagtagga gcgcgtgttc
23761 gaaaaatgca ggatgcgggc attatcgagg gctatacgct gcgctggaac cctgagaaaa
23821 ttggacagac gattcatgcc tttattacgg tttttttgaa ttcaggaaca actcacagtg
23881 ctttccaggc ttttgcacgg gagcatccct atattgtcga aatccatcga gtgagtgggg
23941 aaggctgtta ttggatgcgc ttgcgcatgt catcacaggc agagttaaat accatgctgg
24001 atgagcttac gaagtttggt aattacaaat tgagtttttc cattggtgag atttaacgat
24061 aaaagctttc ggcgtttgtg ccgaaagctt ttttatttgg gcgagggcta tgaggtcaac
24121 gttataggct cattttaaaa atatgctgca atccgtgttg tcccatttta gtgttgcctt
24181 catcgataaa gccgcatttg cgatacagac tttgggcagg cagattacgg gcattgacgc
24241 ccagtacaat ctcgcgaata tgtgggaagt gtctggcgat gaactgtggt agtagaagca
24301 gccctttctt agcgaatcct tgtccctgat ctgcgtaatg aatagagaac gcgcgcagca
24361 gcagtgcctc agaagtattc gtatagtccg ctaattcatc cccatcgtat aaaatgaaaa
```

```
24421 atccaacggg tcttctagta gccgtaatca cgacaggatg gcggtgcgga tcttccttgg
24481 ccagtgtcag caattcatcc ggcattccag taaactgctt ctgctcctca ggcaagtaga
24541 aggtatgtaa aatggcatca tgttcacaat ggtaagggac aagctctaca ttagcagttg
24601 aagaagactg tgaggcgaat gtgggattat tcatcattac actccgatct gtatggattt
24661 ttaactttct ttatcatcta gtttgctaaa aaaagaaacc tgaactcgcc ttacttacgt
24721 attacccata cataaactag catagtcaaa aggagacgtc tacggttatg aagcagctta
24781 cagaatatcg ggaaagccag gacgtattgc gtcatgagct gaagaatatt gaaaaatggg
24841 agaaggaaca gaaggacatt tttttctggg aaaaaatcgg cagatggcct tttatgttgc
24901 tggatcggct gacgcccaaa gtcattaaag acaagctgga gcagctgctt aacgagatgg
24961 gcagcttcat tcagaacggc ggcaaatatt tggtgaagga agaaacggtt ctgaacaggc
25021 tgaaaaagac ggcccatgaa catgtcattc agcaaaatgg tgcagatttc gcttcggata
25081 cgcaggacct gcctcgtgcc ggggaaaacg ctgattcgga acatagtgcc gaacctacgt
25141 gggagttgaa gcaggcggct aagctcccgc taacaatcat ggatcggact gctgacgata
25201 tgacgtcggg tcgcattacc tttgccaccg ctcagggagc aacaacgggg attggcggtg
25261 tgtttacgat tgcggcggat attcctcttt tgctcggctt gtcactcaaa gtgcttcagg
25321 aaatagcctt gtgttatggc tttaatcccc atgataagca ggaacgcatt tttattatta
25381 aatgtatgca atttgcttca tctgatattg tgggtaaaaa ggctgtttta gaggagctgg
25441 ctttgttcga tgatccttca cgtcaggcac aggtttttc ccagatgcag ggatggagag
25501 aggtcattaa cacttatcgt gaccagtttg gatggaaaaa gctgcttcaa atggtcccga
25561 ttgcgggtat tttgtttggc tccattgcga atcgaagcgc tattagtgac gtagctgaag
25621 cagggaaaat gctgtatcgg aagcgtcgga ttttgatgcg tttggcagaa gtagagagtg
25681 aggaattacg ttagttgggt ataaaggtgt gcagcggcca agatggccgc cacacgtttg
25741 cctgttacac ttcggctcca ttttcttccg cccagtggtt cagggtttta tctgtaggaa
25801 gcaggaaggt taagatgccg agcagtggca aaaagccgca caggaacatg atgtgagtga
25861 tgcctacttt gtcgatccag ttgccgagca cgagtgcgcc cacaccacca aggccaaaag
25921 ccaggcccgt aatgagaccg gacacggtcc cgattttacc aggcacaagc atctgtgcat
25981 atacgactgt gaccgaaaag ctagacagca taataaatcc gataatcgcc agcaggattg
26041 ccgtccagaa ttggttggca taaggtagtg ccagcgcgag tggtgcagct aataccatgg
26101 acagaaaaat catattgcgt ttaccaaaac ggtccgccag cggaccgcca aagaaggtgc
26161 ccagcgctcc agctccgagg aacaggaaaa tgtacagctg tgcatcagcc aatggcattt
26221 taaatacttc catcaggtaa aacgaatagt agcttccgat cgaggtcgag taccacgaac
26281 ggacgaatac cagtagaatg aggacaatca tggcaaacat aacgcggtta cgcacttcgg
26341 gcttcatgcg gcgtgccacc gctttttgc gtgcaatgcc ttctgtacgt aggacgttac
26401 cgtaccagcg ggcgatgaac gattgcacca caattcccgc tgccgcaata cctgtgaacc
26461 agatcgcgcc aaattgtcct aacgggataa aaatccagct cgtgagcatc ggcgccagcg
26521 attggcccgc attgccgcca acttgaaaaa tggattgtgc caatccccga cgcggcccgg
26581 ccgccatatg ggatacgcgt gagccttcag ggtgaaaggc agccgaaccc aggcccacaa
26641 agatgactgc gatcagcacc gctttataat cggcagcata ggcgagcaac agcatgccag
26701 tcagcgtaaa gcccatgccg atgggaagga tagcgggcgt agggcgtttg tctgaaaaca
26761 agccaatcac gggctgcatg attgaagctg taaaattgat cgcgaatgaa atccagccga
26821 tctgcgtgta gttcaggctc atagattcct ttaagatcgg aaagatcgct ggaatgaccg
```

```
26881 attgaatcga gtcattgaac agatggacga aactgatcgc aatcagtatc gcaaaaacgg
26941 tgcctttcac ttgcggcccc gggaggggga gcgattggcc cgtttcattt tttgtagcag
27001 atttactagt catggcaacc tccgaatatg tatgtatatt ttaagatata ggaagcgatc
27061 cgtttgcagc aaatggctga aagtgggata aagccctact ttactcattg ctgcttgccc
27121 attctgctgg atactcttta ttgcacgatc ctttaattca cgatagatta tcttgctctg
27181 agtgcattat atctattata aggtatcctt tacttctcat tttactcgct aatatgcatg
27241 aatgagttca actcatcttc aggaattatt agatttccac agtagaccat gttggagtga
27301 aattcccacg atccatatgc caccaaatgg ttataatcct taactctttt aattaatgcc
27361 caaataccaa gattactttg tgcgtttgca tccaggaaaa tccaatcctt gtattgaaat
27421 agttgaggct ctccaaacat atgaaatgaa tccaaatagc ctgctccatc acgcggtgga
27481 atttcattac atacatcttt taataataga acttcctcca ctctaacatt gatccaagct
27541 cttttctcat ttttggatat gatgtcttca aatctacatt taataaacgc agagaagttg
27601 atctcatcag gatgctcatc ccaaccatta aaatttgaat aagcaggtct gaataaagtc
27661 caaaaggttt ctccaacaga catatgaaaa ggaaaatata cttgtacttt ttttgagtca
27721 cctttgccg gggtttcaa tgctctttcc catatttcag tatttagatc tggagtatca
27781 tgcttgaaat ctgcccagtc catagcatct tcgatggcca caatctcaaa atttttgtct
27841 tttaataggg gtttatctgg tacctgttta cacgatcctg cccatgtggt gaattttatg
27901 cgtttcatcg cagaagatgt ccgtgaaatt atggcggaat tgggcttccg tacgattcag
27961 gaaatgattg gacgtacgga ttgtctggat accgtgaacg ccgattcgca ctggaagaag
28021 aagggcgtag acctgagctt gctgttgcat gttcctgagt tggaagaagg tagtgtacgt
28081 taccgcgtac aacgccaaaa tcacggtctg gaagaaacac tcgacatgca acagctcgta
28141 ccgatggctg cggaagcgct ggagaacgga acgcctgttg agggtacctt acctattaca
28201 aacgtgaatc gggcggtagg aacgattctg ggcagcgagg tcacccgtaa gtatggagcg
28261 gcgggtctgc ctgaggatac gattcgcttt aactttattg gatctggcga taatccacac
28321 cgattagctt gcgagcgata tgtgggataa ccaaacctac aaagccgatg gctccggcag
28381 ccgatacaga tagtccggcc agtaccacta cagtcagcag acccaaaagc cggatacgcc
28441 tgagattcaa cccaaggttt aacgccacct catcgcccat tgaaatcagc gaaatggaac
28501 gtgcaagccc catagaccag gcgatctcat cacagcggtc tagaaaatga cgggactgag
28561 ggtagtggga tgtccgtacc atattcagcc ccagctcctc cttgagcaca tctgcatcac
28621 gacgctgcgc tcttttgggc atagcatagc caacgtacgg ataggactgg tgacggttca
28681 agccaagcag cttgatcttc cgcccgttca agtaaaaacc gtccggctga aattcggcct
28741 cgcggaagcc aaagcgcacg gtcacctgat ccaactgttc tccgcctcgt agcagcgtca
28801 gcttcgcttc atacagcttg ggatcatcta tatcccaaag ccgcaggccg cccagctctt
28861 tcagctccag cgttacttcc gccgccgtca ccgctgccag ctcggtctcc gccttcacgc
28921 tcgctccgtc cagcaattgc agcgcgaccg tcaagtcgtc ctgctgatca ggaagatatg
28981 atgaatcttg ttgactcttc acttgatttt tggctgaccg ctggcagata taccgatgag
29041 tttgaggaga gccttgcaag ctatgtgggt gtacggtttt gctcgcttgt caactcaggt
29101 tcctcagcta acctagtagc ttttatggca cttacctcgc caaggctggg agatcgggcc
29161 attcgtcgtg ggatgaagt tattacagtt gctgctggtt ttcccactac agtatctcca
29221 attattcaat ttggggcagt acccgttttt gtggatgtga acataccgca gtataacatt
29281 gattgtagcc agttggaggc tgcgttatcg gatagaacga aagctgtaat gattgcacat
```

```
29341 acacttggaa atccttttaa tgtttcagaa attcgaaaat tttgcgataa acataagcta
29401 tggctcattg aagataactg tgatgctttg ggaactgaat atgaaatgga cgggcagtgg
29461 agaaagacag gctcttttgg tgatattgct acatccagct tctaccctcc tcatcatatt
29521 acaatggggg aaggtggtgc agtatacacg aagagtggtt tactgcataa aataattcgg
29581 tcatttcgcg attggggtcg tgattgctgg tgtgacccgg gagtagacaa tacttgtggg
29641 cgtcgtttta gtggacaata tggggagctt ccctgtgggt atgaccataa atatgtgtat
29701 tcccattttg gatacaatct gaaggtaacc gatatgcagg cggctattgg ctgcactcag
29761 ttgaaaaaac tcccggcatt cgtggagtcg agaagaaaga attgggagta tctgcactca
29821 ctgcttttag attgtcaaga caaaatcatt ttacccgaac ctgtggaaaa ttcaaaacca
29881 agctggtttg gattttgat tacggtgagg gaaggaataa gcaggaacga acttgttcgt
29941 tatctcgagg agcaagggat tcaaactcgt atgttgttcg caggaaactt attgaaacac
30001 ccttgttttg atgagatgag aaaagaaggc aagggctacc gtgtggtagg ccaccttacg
30061 caaaccaacc gaatcatgga caacaccttc tggattggcg tctatccggg catgaccaat
30121 aagatgatag actacatggc aaacaaaata aaacaatttt tacatgagca tcctgtggat
30181 gttgtataga ccatgagagt tgtaatacga gcgtgtgcct ctcttcttat gggtactggc
30241 catataatga gatgtctcac tctggctgga caactccgtt taagaaaaca tgaagtgata
30301 ttgtctgtag tgtagagacc actagggtaa tatggtcgat tttattagaa gtaaaggcta
30361 agaagtactg cctaaacaag cgggttttac ctacagtgag cgttgtggaa tcgatgagaa
30421 gcagttccta gggggagtcc agtttccgac gagtttcccg attcgattta cggacgagca
30481 gttcaaagag ttcgttaaaa accgtaaacg gaacttccgc tgctttggtg gaaaaccgag
30541 aataatgaac cacgggcaag ccactggata tcgcacagtc aacccctgag cgatatctac
30601 cttcttgttc gtacgcatcc aagcaccagt attgcgctaa atgttataat gtgaactttc
30661 gagcatgatc ttcgtacccg atcagttgag caaccgtttc gacttcttct gaagtcagta
30721 cggattgaag gattggcgta atcgtgatgg tatttttcgt gagatcgtct ctttcttaaa
30781 tcgcttggtg gtacaaacga ttttacagaa aagatggtct tttttactat tttcaagtta
30841 gttattcgac ttgcagggtt taatctgaag gagaacagcc ctggaaagaa gaaggggcaa
30901 gtccaccatc accaagcgtg aacgatgaag gcaaggggcg ctgtcattcc gtgcaatact
30961 gcccatgctg gcaaagaatg ctgtgtttaa agcgctacat ccctaatcca tgaagcgcaa
31021 tcacaatccg ttcaaaaaga aacaatcgat cgtggcgtta tgtggaaagt tatacgtgtg
31081 ctgtacactt taggaacaaa acgtatttgt tatcatgcaa ctggtatttt aggagcggta
31141 cgccgattcc gacttcaact ggctactcaa actattttaa accaatttaa accatttttt
31201 cattcaagtc taaggtttaa caacgacaac tgaagcacgg agagccgcag attcttagac
31261 cataacggca cgaacaagta aaggagcaag aagcgccccc ccttgagagg tagaacgaag
31321 gaatatgaga gctaagaccc cgtgtggctg ggagggtaaa ccgtcaaagg aatgatgtgg
31381 atatccaagg cgcgatcata agtaacatcc acagaatgga aaatgagctc caagctctat
31441 tcctgtcacc ggcatcttaa gaaaatatgt cactcgagcg tgatttgccc gactcttaag
31501 gaagtcgaag ttgaaaactt aagaaaaagg gagcaaatga gagaaaacat taatatatag
31561 tgggagggtt cacagttgag agtccttctc tgtggttcta cgggcatgct tggtaaagca
31621 gtggttaagg aattttccaa gcatggtgtt gagttgtaca cattgtcccg gacaaatgcc
31681 gatatcaatt tagatgttgt tcatgcagaa cttctgacaa aacatttgaa tgaactggaa
31741 ccggatatca ttattaatgc attagcaaac gtaaatttgc aattatgcga agctaaccct
```

```
31801 agcagtgcct atttaataaa tgccagacca gtttcgtttt tagcagatta tgcatcttct
31861 cgaaataagt atcttattca aatttcgaca gaccattttt atgcggatca gaaaaatata
31921 ttacacgatg aaaaagctcc aattacattg ttgaacgagt atgcgagaac aaaatatgca
31981 gcagaatgtt ttgctcttac ggccccgaat agtctagtta ttcgaacaaa tattatcgga
32041 ttcagaaatt tggaagcgag acctacgttt gttgaatggg tgcttcgctc actgcataac
32101 caagaaaata tgaaactctt tgaaaatgtg tacgcttcta gcattgccgt atctcaattt
32161 gcagagatta tgtatgatca aatgttggag gcaaggccat attggcatcg aacggaagcc
32221 actgaatata ttttccgggg gcgttccttt ggagtgttgc cgctaacccc tgagccgggt
32281 acaaaatttt atctatattg actgggtccg taaaatacat cgaatctata aaaccggtgt
32341 cttcttctgt acgctccaaa atagccgcac atgcagcatc tccatagttc gcataagtaa
32401 tctcatcctc cggaccggaa atgagtgaat tataatctga accgataata agcgctgtcc
32461 gcacgtcagg attagatagc atcgaccgac tggcttgatc gacagcgact gtcatccctg
32521 cacaattcgc gttggaatcc atcacggcca catcatgtct ggccccaatc gcctggtgaa
32581 caaacaatgc attcatcgga aacagtgact caggaacttg agtggagaac acgatcatat
32641 ccacgtcacg gcctgtcata ccagcctgtt ccagagcttt ttgcgcagcc gcaattccca
32701 ttgttagacc gttttcgttc ggatcatcaa tgatataacg ctgttttcta cccatatgct
32761 ccatgaagcg tcgaatatct ttgtcccgct cgtcaaaatg gcgaacataa tattcgttat
32821 ccaacgaatg gggcggatgg tacacaccca gtgcttttat tttgatatgg ctcatcaggt
32881 ttcccccttta cacatacgtc agcgcattat tgtacgacaa tatctacgtt gtttaatcct
32941 gcttcacgtg ctactcgacc aagctgcatt ttaataatgg tttgttgaag actgatgcca
33001 aagattactt tggcaaatcc agtttctttg taacgtacaa atgagccctt caatttttcc
33061 acttctcccg gacccaaaag tgccatctgg gtacaatcta ccttgagtgt aaagctggac
33121 gcctgaatgg atgaagtttt ggctacataa tcttttacaa agctatcata ctcctcctgg
33181 ctaaaaactc cattcactgc catatgtacc gttttcatgg cttgatctac ttcgatagtg
33241 taaggatgtt gactcattta gaaagactcc tttgatttca ttttttgtc tattaagagt
33301 atcggaataa taggtagaat ttttaatatt aaatcagata aatcaccact aaaatattac
33361 aaatttgtga atataaaaaa ccgtcactgc tcctagaaga agcaactgac ggctggaaat
33421 ttgcgtatct catatgaaat taattcgtaa acgtatggcg accgagtgag atgacagccg
33481 ggcgactttt actccatttg gctgtggcaa tggccgggtt gtaataaaac aaggacccat
33541 tagtcggatc agcaccgtta agtgcctgtc tggctgcacg gtatgccgtc ttgttgggtt
33601 tcaggcggta ttgaccatcc tgtacagcgg tgaattggcc aggctgcaaa ataacgtccg
33661 ggatcgagct tggaaactcg ccggactgaa cacggtttaa tacgacagcg ccgacagcca
33721 cttgaccttc aaacgattct ccgcgcgcct cagaataaat gatacgagcc agttgatcat
33781 gcaacgccat ttcaatgcgt tcgtagctat atttatcatg catgtagtga ataacattca
33841 aagtgttcat gtacagtttt gccaaccatt ccatcatagg tttgaagcgg ctcaacactt
33901 cttcataatc cagaacttca cctgtaatgg ctgggaattc aggccctact tggacgccag
33961 atttttcatc tttaccaccg ttaattgcat acagcagagc tttagccaag ttggcacgag
34021 caccaaagaa ttgcatttgt ttaccgattt tcattgccga tacacaacaa gcaataccat
34081 aatcgtctcc ataaatagga cgcatgagat ctctggggat tcaagcattt taaaggttga
34141 taaggggctc attactcctg tgtcaaaagg gcagattaag attgttgcca catatgagca
34201 gcaaacggta agtacaacga ttacagtaaa atcgccttac agcaagctgc aattaaatcc
```

```
34261 gtctagtccc attaatttga cgattggcgg agaatcaaag caattgactg ctcaaggtgt
34321 acttactgga ggtgggacgg aaaatgtatc cggtgtcgaa tggacttctt cgaatgcggc
34381 tgttgctacc gttgaagatg gtcttgttcg tccacttgcc aaaggcgtga cttacattaa
34441 ggcaagcaaa gatgaactta aagcacaaac gattgtgtat gtgcgttctt cttctcaggc
34501 gttagtattg agtgcaactg gggccaaatc ggtctttctg ggtagctccc caattcaggt
34561 tagtgcaacc gatgttaacc tgactggcgg caaaacagat gtaaccaaca cggcagaatg
34621 gacttcctcc aatccgctgg ttgccacagt cgagcaagga aaaatcacac caatggatgc
34681 aggcaattcg acgattacag cgtcctatca cggtttgagc aagactctga agattaatgt
34741 acttccaacg gttgaaaaac tggtttcgag caaagccagt ctgaccctcg aaacaggagg
34801 caaaacttct ctcccgagta tatctgctta tttggtcaat ggagcaaaaa aagcagttca
34861 atccgatgta aaatggagct taagcagtga aagtgcagtt aaaatagcaa atggtaagct
34921 agttgcagtg aacccaggca gtgcaacatt gacagcaacg gtaggtgcac tcaaactgga
34981 tattccagtg actgtgcaat acaaagtact taaactaacc gcttcagaga aaaaatacgt
35041 attggttgca ggtcaagaag caagtattcc gacagtaaaa gcacacatgg ctggtggcgg
35101 aatccttgat gttaccaatc aggtaagctg ggtaggtaca acagccgcag ttacagtagc
35161 aaatggtaaa gtaaaagcgg tgaatggcgg taactctgga atcaaggcta tgtatatgaa
35221 caaatatgtg aaagtacctg tcattgtgga aggagccatt tctacattga caccaagctt
35281 ttcgagtgct gatatgaatt tgaaaggaag caagtccatc aaagtcatcg gcatatatac
35341 ggatggtaaa aaagctacgc tgagcagtaa agtgaaatgg acaacatcga atgcttccgt
35401 agcgatcgta aaaggctctt ccattaaagc cgttggaatt ggaaacgcaa cgattacggg
35461 cacatatcag gataagtcgt ttaacgttga gatcaaggtc acacctaagc tgctcaaact
35521 ggtattgagc aataaaaacc tgaaattgcc taaaggctcc tctcaagtac ttagtgtgaa
35581 tgcagtttat gattcaggcg caactaccaa tgtgacaagc tcagcggtat ggacatcttc
35641 caaaccgtcg atcgtacagg ttactggcgg tcaggtgaag gcgattgata ccggaagttc
35701 cagtatcaaa gtagtctatg gtggtaagac tgtagcgaca tcggccagag tgttaaaata
35761 gtctacgcat attgctacat ccacttatac aaaataaccc ccggtccaaa gacaggggt
35821 tattttcagc ttattctacg ccgtcctgat aatgattttt ccagtgctta tcgcctttgt
35881 gtgcgaagtc ttcaactgtc ttttccaagg tagctaccag ttgggcttta atctcagaaa
35941 ttgtgtctcg gaacagtctt tcttccgtac cattgatttt ttcaccatat agagcctgag
36001 cggctataaa ttgacgatgt tgctccacaa aactgtcaat cgaatgcaat gcatgctcca
36061 cgtattggga aatgctatga tgctcgttag gcagttgcaa ttttaattca tgaactctgt
36121 ttaataatgc catgatgatc aaaagtgtta accgctttga ttagtccaat agcaccgatg
36181 gatacgagat catgtcgcaa tttcttcgaa tgaactgacg ataagtgcgg aacaaaccag
36241 tgctgcaagt gagcaggtgt cccaggctgt gcaagaaata gccagtggtg cagaaaagca
36301 gactacaggg ctggaaaata attcggtggc cctcgatgaa attgcacagg ggatcaccca
36361 aatcgctgaa cgctcgattt cggtagctga gttggccaaa cgatc
```

Appendix C: The nucleotide sequence of the fusaricidin biosynthetic gene cluster and the flanking regions from *P. polymyxa* PKB1

Sequence 48010 bp; 12497 A; 10395 C; 13002 G; 12116 T

```
   1 ggggcatgaa tggtacgtag ttcatcacgt gaaacgtgtg caggaataca ttgcccaagg
  61 ggatgtattc caggttaatt tgtcgattcg tcagcaacgt gcgcttacag ccctccctga
 121 aaatatatat gaatggcttc gtctgctgaa tccttcccct tatatgggta tgctgcgtat
 181 gccagattta cggatcgtta gtggctcgcc agaattgttg gtgaagctgg aggatggacg
 241 tgtgagcgcc cgacctattg caggtacgag aagaaggggg ctaacaccgg atgaagatgc
 301 cgctatggca gcagagcttc tatcaagcga aaaagaacgc gccgagcata ttatgctcgt
 361 cgatctggag cgtaacgata tcggacggat tgccgaatac ggatcggtgc atgtgcctga
 421 actgctgact gtcgaatact attcgcacgt tatgcatctg gtatctcagg tcgagggaaa
 481 attagccccg ggtcgtacag ctatagatgt aattgctgct acctttccag gcggaaccat
 541 taccggggca cccaaggtgc gtaccatgga gattattgaa gaattagagc ccgttcgtcg
 601 gggtccttat acagggtcac ttggatggat tgactataac ggcaatatgg aattaaatat
 661 tataataaga acactagctg ttaaaggcag cactgcctat ttacagacag gtgcggggat
 721 cgttgtcgat tcagacccgt acagagaata tcgagaatgt cggaacaagg ccagagctgt
 781 catgaaggcg gtacaatgtg gtgaagaaga agctgctttg agcagaagcg gtatcacagg
 841 agggtgagtc gtatgatttt ggtgattgat aattatgatt cgtttacgta caatcttgtt
 901 caatatttgg gtgaacttgg cgaagaagta accgtaaaac ggaacgatga gattgatgtg
 961 aaaggcatag aagagctggc accagagcat attttgatct cgccgggtcc atgcacgccg
1021 aatgaggcgg ggatatcact ggatgtgatc agccacttta aaggtcgcat tccgattttt
1081 ggcgtttgcc tcggccacca ggccatcgga caggcttttg gcggcaaggt tatccgggct
1141 gaacgactca tgcacggtaa aacatcaccg attcttcatc acaatacgtc agtatttgaa
1201 ggcttgcctt ctccgttcac agcaacccgg tatcactcat tgctggtgga gcgcgagagt
1261 ctgcctgagt gtctagagat taccgcggaa accgcagaag gtgaaattat gggtttgcgg
1321 cataaagaat ttgccgtaga aggagttcag ttccatccag aatccattat tacggattat
1381 gggcatcaac tattacgcaa tttcctaaaa cgcaaggtag gcgtctgatt catgaaatat
1441 atcggagtca acggcgtcct cacggaagcc gcaaaagccg tgattcacgt aagtgatcac
1501 ggctttttat acggaatggg cttgtttgaa acattccgca cctacaaggg tgtcccattc
1561 ttgctggatc ggcatctgca tcgactgcag gaaggttgcc gaatgctcgg cattcctttt
1621 caaccggatg aggagcagct tacgagacac atccaacatc tgatggtggc gaatggactg
1681 gatgaagcct atgttcgtta caccgtatca gcgggcgaag aggtgctagg attgccgacg
1741 ggggactata cgcgtcctaa tcatattctg ttcgctaaac cgttgccctc taccaataca
1801 caaaccgggc aatccacttc cgcatcggcg cttcagttgt tgcaaactcc cagaaacact
1861 ccagagggtg aggttcgttt aaagtcgctc cattatatga ataatattct tgcgaagcgg
1921 gaacttcagc agtatgccga ggctgtccgt tataaggctg aaggaatgat gcttacggcc
1981 aatggttttc tggcagaagg aatggtcagc aatctgtttt ttgtgcggaa taatacgctg
2041 tataccccctg acttgtctac agggattctg ccaggcatta cgcgtgagtt cattttggag
2101 ctggcgcatt tgcgagacat tccttgtgag caaggtttgt accgctggga cgagctgaaa
2161 caggctgatg aaatttttat gaccaactct atacaagaaa tacggcccgt ggatttgttg
```

```
2221 ttggagccgg gtggagcaac gaaccatgta tcggttccct tgacgcagcc caggtccata
2281 acagctttgc tgttgcacga ttatagacag aaagcaggga tcaagtgaac ataccgaccc
2341 tttatcaacg tagttaccga atgggggatg ccgagcttac gctcggacat tccacacaga
2401 tcatgggtat actgaatgtg accccggatt cattttccga tggaggatta cattatagtc
2461 ctgagattgc tgttttacat gcattgaagc tagtggaaga tggagccgat atcattgaca
2521 ttggtgggga gtctacccgt ccgggtcatg atccggtagg tgtggaagag gaattggctc
2581 gggtcatacc ggtcgtcgag gccattcatc gtatcgcacc gcacattccg ttatctgtag
2641 atacttacaa agccgaagta gccagaaagg cactcgaggc tggcgcacat atcattaatg
2701 atgtgtgggg gtttaaggca gaccccgata tggcacgtgt agcttcggaa ttcgattgcc
2761 cggtcatttt gatgcataat cgtcatgatc gtaattatag tgatttggtt tctgatatgc
2821 tgtctgattt aaaagagagt atccgattag cgctggatgc gggggtacgt gaggaacaga
2881 tcattttaga cccaggcatc ggttttgcca aggattatga tgaaaatatc catgccatga
2941 cttctctcga tacattgagt cagctaggct tcccactgtt attggctacg tcacgcaaaa
3001 gatttattcg tacagcactc gatttgcctg tgaatgatgt ggtggagggc acggcagcta
3061 ctgttgcttt tggcattgct caagggtgcc agattgtgcg tgtacacgat gtggcacaaa
3121 tcaagcgcac tgtacgtatg tgtgatgcga tggtgtacgg gtctaaaaat gcactgagtt
3181 atgacgggta aagcccgctc cacgaatgat ttgatcctac gatcgctgtt gcaacgggat
3241 tctttaaatt ggtaaaaaat gtaattgtaa gaatcccgtt gcaaaggcga acgctccgct
3301 tcttcggatt caaatccttc attacgctac caccagcata aagtcacttg attagaaaaa
3361 tctaataaaa aacaggggc ctacccacga tggataaaat ggttttgcat cgtatggagt
3421 attacggata tcatggcgtt tttccagagg agcgcaagct gggacagcgt ttctatattg
3481 atttggaact ggagcttgat ctgcgtgaag cgggtgagca ggatgctttg gacaaaacgg
3541 ttaattacgc cgaggtacat tacaccgtta aagatattgt agaaaaggaa tcttaccaat
3601 tgattgaagc tttgggcgaa cgtattgcat cttccctact ggacacttat actagtgtca
3661 atgcactgac tgtcaaagtc acgaagccgc atcggccgtt tgatattcat tttcaaggcg
3721 tgacggttga gctttatcgc gcaagaaagt aggaacgata catgaatgca cattcgacct
3781 ctgagggatc agaggcttat attgctttag ggctaattt gggggagcgc gagcataccc
3841 tgtatgaagc cattacggcg ctggacgaac atccgggtat acgggtgctg cggtgctcca
3901 gcctgtacga aacagagcct gttgggtatc tggatcaacc ttcttttttg aacatgacgg
3961 tagccgtatc tacaacactt gctccagagg agcttctggt atatatgctg gaggtggaga
4021 gtcgtctggg gagagtaaga catattccta atggtccacg aaccattgat ctggacttgt
4081 tatggatggg agcggcgcaa ctggctacac cggatttgga gcttcctcac cctcgtatgt
4141 tggagcgtgc ttttgtgctg gttcctttgg ctgacatcgt tccgaaccaa gatccgtccg
4201 gtctgtacag tatcgtgcat actgcattga attctgtgga tggaaaggac ggcatacagc
4261 tttggaaaac atgcagctgg cgcaacgact ccgcgccttt cggaagctaa aagggttcac
4321 acagcaggag ttggcagaac gtactggtat atctttgact gtattaggtg ctgtggagcg
4381 aggcaatcgt acagtagatc cagacatgct gaacatcatc gcccaaacgc tagagattga
4441 agtccgggaa ttgacggatt gacaataagt atacgatata aaagaaggaa taaaggggtg
4501 aacagatatg ttgaaaatcg gcgatattga aatgaaaaac caagtcgtct tggcacccat
4561 ggcaggcgtg tgcaatcctg cttttcgctt gatcgccaag gaattcggta caggcttggt
4621 atgtgcagag atggttagtg ataaggccat cattcacggc aacaagcgta cacgtgagat
```

```
4681 gctgtttgta gatgagcgcg agaagccgct tagcttgcaa atttttggag gcgaccgcaa
4741 atctttggtg gaagcggcga aagtcgtaaa ccaggagacg aatgcagata ttattgacat
4801 caatatgggc tgcccggttc ccaaagtgac caaatgtgat gcaggcgctc gttggctgct
4861 tgatccaaac aagatttatg agatggtatc ggcggtagtg gaatcggtaa gcaagccggt
4921 taccgtcaaa atgcgtattg gttgggatag cgagcatatt tatgccgtag agaacgctcg
4981 cgctgttgaa cgtgccggag gtcaggccgt tagcgtacat ggccgtacac gggagcaact
5041 gtatacggga catgcgaatt gggatattat aaaagaagtt aaagaagcag tttccattcc
5101 tgtaattgga aacggtgatg tggcatcacc ggaagatgcc cgccgtatgc tggatttgac
5161 caattgtgat ggcgttatga tcggacgtgg tgctctgggg aatccgtgga tgctctaccg
5221 tacgattgag tacctgcaaa cgggcgagct gttgcccgat ccttcaccgg aagaaaaaat
5281 ccgcattgcc attctgcata tggatcgtct ggttgcattg aagggtgaag cagtagctgt
5341 ccgtgaaatg cgtaagcatc tggcttggta cctcaaaggg tttaaaggtt cggcacgtat
5401 taaggatgtc attatggaag aaacaaagcg cgacgatatg gtacgcattt tggatcaatt
5461 tgtatctcaa ttgcacatgg tagataacga tgagccttct acaagcgctg cttctcctgc
5521 tgctgttgcg caataaggaa gctgttccgg cctgtgcacg cgattgactt ttccaattgc
5581 ttgccctata atcgttcagt ataaaatctt gatgtatgct acaaattgtg tgcagctctc
5641 aagacaacag caggaaatgt tctcattccc tcagagtggc atattatgtt ttgaaagcgt
5701 gcattccgat gcacttttaa acttttccca tgacaggaga atcggttaag atgagcgata
5761 aagaagtcat tctgacacag gaaggtctta gaaagctcga agatgagctt gagaacttaa
5821 agtctgtcaa acgccgcgaa gtggctgaac ggattaaggt agccatcgga tacggagata
5881 ttagtgaaaa ctccgaatat gaggatgcga agaatgagca ggcctttatt ga[illegible]ggccgtg
5941 taattacgtt ggaaaagatg ctacgcaatg cccgcattat caatagtgat gaaattgaca
6001 ccgatgttgt gggagtaggc gcaactgtga ccgttgaaga tttggaattc ggagatatta
6061 cagagtatgc tatcgtaggt tccgccgaat ccgatccgct acagaataag atttcgaatg
6121 aaagtccggt tggcaaggca attctcggta agaaaaaggg aactgttgtg gatgtcaccg
6181 ttccagcagg cgttattcaa tataaaatcg tggatatcaa gaaataagac acggactttg
6241 gaatttaatg taaaagcttc cttgtggaag cttttttct ctacacccgg acggaccccg
6301 gcccgtctat tcatattgaa ggagctgatt tagcaccatg tcggatgaaa ctactaacca
6361 ggaatcacaa gaaaatcgcg aggaattaag tgaactattg caaattcgcc gtgccaaatt
6421 ggacgagctt cgctcattgg gcattgatcc ttttggacgc aaatacatac gcacggctga
6481 agcgggcgta ctactgagca agtatgacag tttgactaag gaagaactgg aagaaaagaa
6541 tattgaagtt agcatagcag gccgtattat ggctaaacgt gtgatgggga aagccagctt
6601 tgctcatata caggatctta gtggacgtat tcaaatttat gttcgtcaag acacggtacc
6661 tgaatccaaa tataaagcat ttggtatttt ggatctgggc gatattgttg gtgttaaggg
6721 cgtgctcttt aaaaccaaaa caggcgaaac gacaatcaaa gtgctggata ttgaggtctt
6781 gtccaaatcc ttgtatccat taccagagaa atatcatggc ctgaaggatg ttgaattgcg
6841 ttaccgtcag cggtatgtcg atttggttat gaaccctgaa gtacaaaaaa cgttcattac
6901 acggtcccgt attattcaat ccatgcgtcg ttatctcgat tcgctcggct atttggaagt
6961 tgaaactcca acattgcatt ctattgctgg cggggcggca gcgaagccat ttgttacgca
7021 ccataatgca ttggacatgc agctttatat gcgtattgct attgaactcc atttgaaacg
7081 actaattgtc ggtggaatgg agaaagtata tgagatcggt cgtgtttac[illegible] gtaatgaagg
```

```
7141 tatttcgacc cgccataacc ctgagtttac gatgattgag ctgtatgaag cgtatgctga
7201 ttataaagat atcatgcaat tgacggaaaa catggttgca catattgcgc aggaagtact
7261 gggcacacag cacattcaat atcaaggcca ggaagtggat ctgacgccgc aatggcgcag
7321 agtctccatg gttgatgctg ttaaagaagt aacaggcgta gactttagcg tgcagatgag
7381 taatgaagaa gctcatcgcc ttgcgaaaga gcataaagtt tcggttgaac cgcatatgac
7441 atttggccat atcctgaatg cattctttga gcagttcgta gaggaaacgc ttattcaacc
7501 gacctttatt actggacatc cggttgaaat ctcaccgttg gccaagaaaa acgaacaaga
7561 tccgcggttt acggatcgct tcgagttgtt cgtggttgct cgtgagcatg cgaatgcgtt
7621 cactgagctg aatgatccaa tcgatcagcg ccaacgtttt gaagcacaat tgctggaacg
7681 tgagcacggt aacgacgagg cgcatgaaat ggatgatgac tttatccgcg ccttggaata
7741 cgggatgcca ccaacaggtg gactgggcat cggcatcgat cgtttggtta tgttgctgac
7801 agatgcgcct tccattcgcg atgtgctgct tttcccgcat atgcgtaacc aacagtagaa
7861 taacataaca ttatgaatat agaggatcag ccgggcttcc cgtgtcttgt tcctcttttt
7921 cataagaacg gatggagaga atactctaat gaaacgcagc gtcaagctta gtcggttcgc
7981 atcgcctcct tttatattgg tcatcggata tttattgatt attgcattgg gtactttgct
8041 actgatgctt cctgttgcca atcatagtgg tcatgctccg cattggattg attcgctttt
8101 tacgtccgtc tcagccgctt gtgtcacagg actggtcgtt gttgacgtcg attcaacgta
8161 tacgatgttt ggcgaaaccg tcattttggt gctgatgcag ctaggcggtc tggggtttat
8221 gaccttagcg accttgtttg cgctgttgtt ggggcgacga ttatcgttaa aggatcgatt
8281 gctgcttaag gaggctatta atgccgatag catggaaggg atcgtgcgta ttattcgcaa
8341 ggtgctgatc ttttctttta ccattgaggc gatagctgct gtcgtcttgg ctttgcggtg
8401 gatgcgagag atgcctgtcg gacaagcgat ctattatgga atattccatt ccgtatccct
8461 ctttaataac gggggattcg atctgtttgg gactagcttc caacagtacg ctggagattt
8521 tgttttcaat atgattgcat ctgtgctcgt tatatctgga ggcttgggtt tccttgtact
8581 tagtgatttg ttcgattttc gccgtacccg aagactgtct ctacactcca aaatggtgct
8641 tagcgtgtcg ggagtgctca ttgcgattgg tgcgctggtt atatttattt ttgagtttac
8701 caatgcacga accttgggcg cgctgaactg ggaaggcaaa gtatatggag ctgtctttca
8761 atcggtctcc acgagatcat cgggtacgag cacagtcgat atcagcgggt taaggcaagc
8821 aagtcagttt tttattatgt tgctgatgat tatcggtgct tctccaggct ccacgggagg
8881 aggaattaaa acgactacgt ttctgcttat ggtaggtgcg ctaatcgccg tcatgcgagg
8941 acagaaggaa attgtgtttt tccgccatag ggtccccaaa gatatcatga tgcgggcgct
9001 gaccattatt gtactagccc tatttatatt ctcgattgtg gttatactgc tcttaatgac
9061 agaggatgcc ccatttctct cccttgcatt tgaggctgct tccgcaatcg gaacggtagg
9121 cttgtccgta ggagtaacgc caggattgtc tgatgcaggc aagctgatta tttgtgtgac
9181 catgtttatt ggtcggattg gtcctcttac gattgcctat gccattcgtc cacgctcgac
9241 caagacattg taccgtcgtc ccgaagggca tattgtaatt ggataatgta tctataaaaa
9301 ggatgtaccc gtgctgatta gcataggtac atctttttc atttcagatc atcatcattc
9361 tttatgggga gcatatcatg taaaatgtcc attttgagaa caaaacgaaa gggtctgaac
9421 agatactgtt tttcgcagaa aaaacggcta tttgtttcct tttatgtaat tatagacaaa
9481 atactaggaa tattatgtat aatattgtat gatagtgggc atgaaagggg gatagttgag
9541 gtaaatacca atttctaatt tgaaaggaat catctattat gtatttattt attattagtg
```

9601 ctgaagtttt attttgggtt tttgttgttg caggcttgat cgcaaggtat atgtttggct
9661 tgaagaagtt aggcggtctt cttttactat ttacccccat tattgatatc ctgttactga
9721 ttgcggtgtt catcacggtg aaaagtggca tggagattac gactgctact ggcttggcgg
9781 cctgttattt tggtattaca gtagcatttg gtcataggtt aattaagtgg gcggatgtgc
9841 ggttctccca ttggtttggt aagggaccca agcctgaacg caagtatggt gctgctcatg
9901 caaaagagga acggataggt tggttgcttc accttttggg ctgggctata ggtaatgctt
9961 tgttgctcgc catcattgtg tatgttgatg atcctcagcg tacggtgtct ttggagggca
10021 ttatgcaaac atgggcaatc gttctggcaa tcgactttgt tgtctccttt agctatactt
10081 tagctcccaa aaaacaaaag cgtagtctat aaaggttaag acttacactt agatgatatc
10141 tggctggtca ctcaggatgt tcaaatcgga agagtcattg ttggtgaaaa tgaagtggtt
10201 acttccggta ggtgggtgat tctgtcatcg gaatgtaaat atttacaaaa atatactgat
10261 tttatgttta actcgctttg acagtcacct tgtttttgcc tcgctacctt ttttcacata
10321 catgtattcc tttcgatttg aacacagacc aacagaagca acacaactat ttcttgtgag
10381 atgacagaca gcatacacga tgtgttttct agaccttagt aagaaggatc gttgtaagac
10441 gaatttatgg gtctagtatg gaggcacttg gaatttgaaa aaattgggct cgagattata
10501 gctaaaatca actgcggagt agtgttctac ccattcacac aaattttaaa ataaattttg
10561 tgatttgcct aaaacgtcgc ctatccaagt ttggcaatta acgcgataat attccaggaa
10621 agattacata tgtctataaa tgtagacata cgcaattaag ttgttatttt gtttaaaaat
10681 gtaaaaaaga aaggtggttt ctgtggcagt actactgagc ggtaagaaca ttttgattat
10741 gggcgtggct aataacagga gtattgcctg ggggacagct caatctttag ctgctgaagg
10801 agcacagctc atttttactt acgaaaatga gagggttgca gagcgggtac taaagctcgt
10861 tgaaaccatt ccgggagcac tggtctgtcc ttgtaatgta caaagcaatg aagagcttga
10921 tgcattggcg gattttctaa aagaaaaagt aggcgtattg catggttatt ttcacagtgt
10981 cgcttatgct aaggctgagg atttggatgg cttgttcgtt gacacttctc gcgaggggtt
11041 cgctgttgca caagaaatca gtgcctattc tctggctgct gtatcccgtc gagtataccc
11101 actgatgaca gagggtggaa gtatcgtcac catgacgtat attggcgcag aacgagcgat
11161 caaaaattac aacgttatgg gtgtagcgaa ggcttcactg gaagccagca tgcgctatct
11221 gtcgaatgat ctgggcacat atggcattcg tgtaaatgcg atttccgctg gtccgatacg
11281 gaccttggct gcctatggca ttaaagattt taacacaatg ctcaaacata ttgaagaaac
11341 tacgccccta cgcaagacgg ttgatacgtc ggaagtcgga gatacggcgt tattcctatt
11401 ttcccattta tcgagaggca ttacgggtga agttatccat gtagatgctg gataccatat
11461 tacggggatg taaggaggga acgttgtgct tattacgcat acggattgcc gtattatctc
11521 acctgaagga gttatgacta gggagcgtct ggagcgagac gttacgatct tctcccggat
11581 gttgacgctc agaggaattc ccgaagacgg tagagtgatt ctgaaagcgg ccaattccta
11641 ttggtttgtt gtgtctctgt tctccctgtg caatattgct gtttcggtga ctgtggtgga
11701 cgagcagact gtgctcgatg aggtggccca aatctatgat gaagcgggtg cctgttgcat
11761 ccttacggat tccgatttag aacttcctca tgctctcgtt attctgattc aagatctgat
11821 tcaagaagaa gcagtacacc cggattgggt gggagagcca aagcttgaag ccctcgactt
11881 tagcaagtgg tgcgagaggt cagatgcgct cattctgtat tcttcaggta cgacgggcaa
11941 acccaaggga atcgtgaaag ctggttcagc ttttatggag aatatcaggc attccataca
12001 cgccatgaac tacctaccgt ctgaccatat gctccccgtt gtgccttttt ctcattttta

```
12061 tggaatctca cttatcttct cctggtggtt aacttcatgt tctctcatca tctgcaatcc
12121 aaaaaacctc tggagcgtca ttgcaagtat tacgaaagac cgggcgacgg tcgtagacgc
12181 taacccctct gctttttaca ccttgttgcg tatgctaaac cgaaaacccg aacaactgga
12241 aatggtcaag gaagctccgg tgcgcatgtg gtgcgtaggc ggttccccac ttacccagga
12301 ccttgaagag aaatttaaca acatatttgg acagcccctg ctgaatggtt acggcctttc
12361 tgaactgggc aacgttacgc tggggacgct agagtgtccg aagggttgtg gcctaccatt
12421 acctggtgtg gatctgaaaa tactagattc tacaggcaac cagcaggcgg acggtatcgt
12481 aggcgaagtg tggatacgca gtgctggatg catggaaggt tatctaaatc gtccagacct
12541 gacgcaatcc gttctacagg acggctggtt caagacgggg gacctgggct atctggatga
12601 cggaatgctg tatgtcattg gtcgcagcgg taagacggtc aaccggatgg gctacatggt
12661 atctccggtc tatattgagg accgaattgg ttcgctaggc taccgatcct gcgtgattac
12721 cttggaggat gaggctaaag ggacattgct tgtcgcattc atcgaaagtg attcctcaca
12781 ggtcgtgcct gtgctccgca aagagatgag tcaggtactc ccttcataca tgtttcccga
12841 tcttttgctt ccgctcgatc atttcctttt aaatcgcaat ggaaaagtag atcgtctgga
12901 gatggagcgt gttgctctag agaaagctgc ttctcgcaaa gcataacaag atatatacca
12961 aaactgcata caggaaaggg agaatgtgcc atgagtataa atccagaact ggcagaaaac
13021 acactgtttc tgaaaaggca tgaatctatt ctgaaatgcc tcaattattt gattgaaaat
13081 cggcaggaag tgatggatat tctcacacaa ttttcatcct atcgggctgc caatgcagaa
13141 attgactcca cgatcctcac acttcaaggt gctctgcaag aagtgcagac ctatcagcca
13201 agctggcaac gatctatggc tgtatttatg ccttccaacg tcattttgta ctcctatgca
13261 ttatacctct tgattccgtc cctgtacgtt gaaaatatag actttcgccc ttcttctcac
13321 gtaaacgagt acgtcaacat gctgcatgaa aagcttcagg ctgtacatgg attacccatc
13381 tatattcgta aagttagcca gaaggtgttc atggagaatt ccgtgatgcc agctgatatt
13441 gtggtgttca cgggaagcta taccaatgca gagaagatca aaaaacagat tcgaaaagat
13501 cagctctaca tttttacgg acaaggacta aatcctttca tcatcggacc tgatgcagat
13561 ctggagctgg cggttaccga tgtgatccga atgagattgt tcaactctgg acaagactgc
13621 ctaggccccg acatcattct ggtacatcaa gatgtgagtg agcatttcaa agatctgctg
13681 attcaacagc ttgaccaact cgtgtttggt gccaacgatg atccgaacgc aaactatacc
13741 cctatattct ataaagacac gctatattcg gtttccgaat atttcaatac gaatgataag
13801 tttattattt acggcggcgg cattgatttc cgtacaaaaa agatggagcc aacggttgta
13861 tatagtgaat tggatcagaa tctggagatt attgagtatt tttcgccagt cttcaatgta
13921 gttagctacc aagacgacga gcaattgatc aagcgtattt cctccagtta tttcagtgaa
13981 cgtgcaatgg gttgcagctt atacggatcg gagcacctga ctgacgttct gcgaaagaag
14041 cacacgctga ccattaatca gacgcttgag gatgtggatc aaggcaataa gccttttggt
14101 ggctacggca cgatgtcgaa ctatattttt tacgatttca agctcatttc gaagccgatc
14161 ttgatttcag aaattgtcgc agaatatttg tcggaaaaga ggagtttggt atgactgcga
14221 gcttctttcc ctgggatgcc atcattggct acttaaaatc ggctggggtt cagcatctct
14281 ttggattacc aagtgatgat ctcaaaatcg tatcgtcact cccctcatcc ggtatggatt
14341 ttattctgtg taaagatcaa cgtaatgcgg tgttcatggc agcaggttat gcgctcacca
14401 ccaaccagat gtccgtgtgt gttgtgggga aaggccctgc gctcagtaat acactcacag
14461 ggctactgga agccaaaaac ctgggcgctc cgctcctgct tctggcgctt ggcactggcg
```

```
14521 gagacaagct gggtacccga tcctttcagg aagcggatca aatctcactc gtgaagccgc
14581 tcgtaaaatg ggcttatcgt gtcgagcatg tggatcggct tgtatgggcg atggagcggg
14641 ctgccttttt ggctatcaat ggagcaccgg gtcctgtcta tatcgaactc cctgaaaacc
14701 tgatggacca gcctgtgcca acagggcttc agtttacccc accagaaaag ctgtcgtgct
14761 ctccgtctcg tcgggagttg gatgcggcgt gggggattat tgagaacgcc cgcaggccac
14821 tcattctagt cggaggcgga atgaagtcag gtacaggtga tgtgattgaa cggtttgcta
14881 accaacatgg agcggccttg ttcgcaaccg cctcgggacg gggcgtagtt gatgaagacc
14941 acgaactctt ctgtggtgtt gcgggactct atacagatct tagtctaaga caaatctgga
15001 aggaatctga tgtggtaatc acactgggta gtcgattgga agaaacggct actttcgagt
15061 gggatatttt gctacaggat acacctttga ttcaagtcaa tgtaaatctg gaggattttg
15121 cacacgaata caggggtatc aaggtgctgg gtgatggata tgcagcggtg gagcagtggc
15181 ttgaacggcg cggccaggag cctgatggag tgtggcttga aacgattagg agctgtaaac
15241 aggaggcttt ttctcacagg gctatttatt tggaagagct gaaaaagacc cctggaatcc
15301 acgttcctga gctgctggac atgattcaga cggagttccc tgaaaatctg gtgcttcttc
15361 aagagaatgg attgcaggat atgtggtcct atttttatcc atattttcgc tttgcccaag
15421 gctctctttc gattgttcct agtgatcaaa cgagtcttgg attcggtgcg gctgctgctc
15481 tgggagcaag cgtggcaaca caacgtccag tagtggcact tgtaggtgac ggtgcgttca
15541 atctctttg ttcggatttt atgacggcgg tggagtatca aataccgtta atctatctcg
15601 tacttaataa tggtgggtat ggttggttac agaatcagat gaattataag catctcacag
15661 ggcagtccct cccgtttgtt tctgaggcgg ccaagacagg aattcaaatt cctcagtacg
15721 agtttgtgga aagtcttgca attcggagta agggagaggc acgggaacag cttcagcttg
15781 cttggcaaaa gtatcaggag aacaagcttg tgatcgtcga ggtatatgct gatcttgcgg
15841 atgtgcatga aaaaatcagt catgtgtatg gcgactttcc tctctacgag caacaggctt
15901 ctaagaatta atttcaacga aggacgtgga tacagtgaga agacgtgttg tagtgactgg
15961 acagggagtc gtaacgccga tcgggcttaa tcttacggat ttctggaatt cactgttgga
16021 agggaaaagt ggaatcggac ctgtcactgc tttcgataca gcggacattc cgacaaagat
16081 cggtgcacag gttacgggct ttgacccatt gcaatatatg agcaagaagg aagctaataa
16141 agtcgcacag ttcactcagt ttgctttggc ggctgcccgg caggcagtgg agcaatcaaa
16201 attagtgatt gatgagcgga atgcaagtcg cgtgggagtt attgtcggct ctggagcagg
16261 cggcctcgac gttatagagg agaactttgg gaaattaaca aagcttggtc ccaaaagagt
16321 ttctccttat tttgtttcct cgatgatgat taattcagca cctggtgaaa tctccattgc
16381 tattggagcc aaggggcctt cttttgctgt ggttacggct tgtgccacag gaagtaatgc
16441 cattggtgag gcattacgta cgatccagta cggcacggct gatgtcgtgg ttgccggtgg
16501 agcagaagcc aattttagtc aacttgatct ggcatctttt gcgaatatcc atgctctctc
16561 cagacgtaat gaagagccgc agaaggctag tcgtccgttt gatgtggatc gcgatggatt
16621 cgtcatcggt ggaggagcag gagtcttggt gctggaggag ttggagtatg ctttgaaccg
16681 cggagccact attctggccg aagttgtggg ttacggatgc acaactgatg cgtatcatat
16741 cacggctcca gatccctcag gctccggccc ggcagaagcg atcaatatcg cactgcagga
16801 tggcggctta acaccgcagg aaatcgacta cgtcaatgcc cacggtacga gtacgccatt
16861 taatgatcgt atggagatta atgctattca aaaagtattc ggggagcatg cccccgttt
16921 ggcgatatcg tctattaaat ccatgactgg tcatttaatg ggcggtgcag ggcagtcga
```

```
16981 attaattgcg actgtgcaga gtatgattca cagcaaagta ccaccgacac tcaattgcga
17041 caatcctgag gctcctgaat tgaactttgt gccgcatgta tatcaggaac gcgaggtccg
17101 cgcggcgatt agcaattcat ttggtttcgg agggcacaac gtctgccttg ctgttcgcaa
17161 gtgggaaggg gaataagcct gtgagcgaag agctaatcca tatcccggat gtgcttccgc
17221 accggtaccc atttcttctg ttggatggcg tgaccgggta cgaaacgtcg caatgggcga
17281 agggatacaa gcttgtcacc tggagtgagt ggttcattac tgacactaac ccatacatgc
17341 ctcctatgtt gattgtggag tcgcttgcac aattagcggc attcgctgcg attgataaaa
17401 aaggcatttc atacctgact aacctcaatg gaattcgatt ccacagcttg gccaagcctg
17461 gagaccgaat tgatctctat tttgaagtga tcaagcgtcg gcggggctat attctcggaa
17521 gcggtcgggc ttccgtaggg gatcgattgg tggttgaggc ggaagaaatt gtttcactgg
17581 atcaatgacg gggtggaagc gtacgttaaa ctgtcctttt gtcctcctga ctgccaggtg
17641 cagcgggagg atattcacat catggaatag gagtgaatga gatgaatgac atgcagttat
17701 atgatttaac aaatgcgcag aagcgtatat ggtataccga attactctac ccagatacgt
17761 cagtgtcaca gctttccggt acagctaaga tgaaggggca tatcaatatt gctgccttta
17821 tgcagtccat taatttgatt atcaaacagt atgatgcgtt ccgcatccgt attacctcag
17881 tggatggagt gcctcagcag tacgtcgttc cttatgaaga gagacagttg gagtgcctgg
17941 atcttagtca ctatgaaagt gtatctgagg tggaagcctt acttgagcaa cacaaaagaa
18001 aacccttgcc cctgctggat tctgagctct tccagttttt aattgtgaag attagcgagg
18061 aagagtattg gattaatatc aagatgcacc atattatttc tgacgggatc tcaatggtgg
18121 tctatggcaa tcagctgaca gcattttaca tggagttaat tcaaggaaat gaaccgaagc
18181 tgggcgacga ttgctcgtat attcaatata ttgcagatga gaatgcatac gaactttctg
18241 acagatacca aaaggataag gcttactggc tagataaatt ttccgatttg cctgagctta
18301 cgggttggaa gtcatataat ccgttatctt taagcaccca cgccgttcgg gagcatttta
18361 ccgtaccaga agtgctatat cacgagctgc aagcattttg ccaacagaat aggatttctt
18421 tgttccagtt cttcatgggt gcgatgtata tctacataca caaaatgacg aatcagccgg
18481 atgtggtgat tggcacttcg ttcgctaacc gggggaacaa aaaagagaag caaaagatag
18541 gtatgttcgt cagcaccgct gctgccagaa catacgtcaa aaaggatata gatgtgttga
18601 gcttcctgca ggatgtagcc agagatcaga tgtcagtcct gcggcatcag aaatatccat
18661 acaatcagtt aattcaggat cttagagaaa tgcatgggaa caaggatatt cagcggcttt
18721 ttggcgtttc aatggaatat cgtcttatca attgggttga tttggatgat gtgcgtattt
18781 tgacagatta tgatttctgc ggggacgaag tgaacgattt cgtgtttcat atcgtggaga
18841 tcctggatga aggcgaactg gtactggatg tcgattatcg gacagagctg tttgaacgca
18901 gtgaagttaa ggacatggtt tcccagttgc ttacgatcgc cgagcagatc attcattcac
18961 ctcagctttc tatcgcagag gtaagcttat taggtgaacc agaagagcaa tccattttgg
19021 ctctttcgga aggcgccgca gtcgattatc cacgagagaa gaccattcat ggcttattcg
19081 aggaacaagc cgagcgcacg ccagatcacg tagccgttca gatggacgag cagagcatta
19141 catacctagc tctaaacgag caggctaacc agcttgcgag atatttgcgc tccgagggag
19201 taggggcaga tacgctcgta gggattatgg ctgaccgttc cttggagatg gtcatcggga
19261 tgttggccat tttgaaagca ggtggtgcct atgtaccaat tgaccccgat tatcccgaag
19321 agcgtatcca ctatatgctg gaggattcag gtgtaagtct gttgctcacc caaagtcatc
19381 tatgggagag caccactttt gacggaaagc ttgtgagtct ggacgaagct gcaacgtata
```

```
19441 caggagatgc ttccaatctg gagagtattt cgggaccaag ccatctggcc tatgttatct
19501 acacgtcggg tacgactggc aagccgaagg gcacgctgat tgaacacaaa aatgtagttc
19561 gactgctctt taacgataaa aatttatttg acttcaactc tcaggatacg tggacgctgt
19621 tccattcgtt ctgcttcgat ttctccgttt gggagatgta cggagcgctt ctttacggag
19681 ggaaattggt gattgttcca tctctcacag ccaagagccc agcagctttc ctggagttgt
19741 tgaaagacaa ccaagtcacc atcttaaatc agacgccgac gtatttttat caggtgctac
19801 aggaagagtt aatgcactct tcgacagagc ttggccttag aaaaatcatt tttggtggag
19861 aggctttaag tccatctctt ctgaggaact ggcgggtcaa gtatcctgat gtgcagctga
19921 ttaatatgta cggaattacg gaaacaacag ttcatgtcac ctataaggaa atcacggaac
19981 atgagattga agcggggaaa agcaatattg gcagaacgat ccccacactt agcgcttaca
20041 ttctcgatga gcaaagacgc ctgcagcctg ttggggttcc aggagagcta tacattgcgg
20101 gagacggtct tgcccgtggg tatttgaatc ggccggattt gacgtctgag aaattcgttg
20161 agcatccgta tcgggtggga gagcggctgt accgaactgg ggatcttgct cgttggttgc
20221 ctgatggcaa tattgaatat ttggggcgga tcgaccatca ggtcaaaatt cgcggctacc
20281 gaattgagct tggcgaggta gaagcccaaa ttctcaaggc tccaagcgta cgagaaacga
20341 ttgttctcgc acgggaagac gaacagggcc aaaaattgct gtgcgcctac tatgtagcct
20401 ccagtgatct ttcgccgggg gaattgcggg ctcagctggc ggcggaactc cccgcttaca
20461 tgattccttc ttattttgtc ccgctggagc aaatgccgct tacaccaaat ggcaaactgg
20521 atcgccgtgc gttgccggct cctgaaagca gcgtacaatc cggtgaggct tatttggctc
20581 cgagaactgc tgtggaagct cagatggtac tcatctggca agatatcttg ggcgttgctc
20641 gtgtcggtgt cagagataat ttctttgaaa ttggtggtca ctctttgcgg gcaacagtgc
20701 tcgtttcacg gattcacaaa gaattgggat gtagcatttc gctgcgtgag gtgtttcagt
20761 cacctacggt cgaatccttg gcgcaacttg tgaaaaaaca cattccgacc ctgtacgaat
20821 ccatcccaca ggcaacggaa agcgaagctt acccagtgtc ctcagcgcaa aagcggttat
20881 acgtgctgag acagatggac gggggagagc tcagctacaa tatgccaggg gtcttcacag
20941 tggatggact gttggatcgc acgcggctgg agtctgcgtt ccaggcactg atccagcgtc
21001 atgaatccct gagaaccggc ttttatatgc aggatggaga gcttgttcag cgtgtgcata
21061 ggaatgtgtc gtttgctttg aactacacag aggctttggt ggaggagacg gatacgctca
21121 ttcacaactt tattcgtgcc tttgatctga gccaggctcc attactgcgt gttagcttgg
21181 tgaagctcca ggaggagcgt catctgttgc tgtttgatat gcatcacatc atttcagatg
21241 gggtttctat tcaaattttg atagaggaac ttactcattt gtatcaagga gaacagctac
21301 cagagctgca catccagtac aaggattatg ccgtatggca acgggaacag tcagagaacc
21361 aatggcaaga tcttgagaaa tattggttgc aatcctttga aggagagttg ccggtattgg
21421 atttgcctac agacttccaa cgaccttcag ttcggagctt cgagggtagc cgaattgatt
21481 ttacattgga tgagtctggc aataaggcga tacaagagct tgcatcccgt acaggtacta
21541 cactgtatat ggtattgctg gccgcttatt cggtactact gcacaaatat acaggacagg
21601 aggacatcgt cgtaggttct ccagtagccg gaagaccgca ggctgagctt gagggcatca
21661 tcggaatgtt tgtcaacaca ctggccttgc gtagctaccc ggcaggagat aaaacctttc
21721 aggattatct tctggaaatc aaggaaacgg cgctcaaggc gtttgagcat caggattacc
21781 cttttgaaaa attggtcgaa aagctgggcg taggacgtga tgtcagccgc aatccgctct
21841 ttgacacctt attggtatta caaaataccg agcaggaaga gcaggatatg gacggagtgc
```

```
21901 actttactcc ttacttgatg gacaccgtca cagccaatt tgacctatcc ctcaatgtag
21961 aggagaaggg atcaaaacta gcctttggcc ttgagtatag tacggcttta tatcggcgtg
22021 aaagtgtaga gcgacttgca acgcacttgc tccggattct gcacgcagtc gcggccaatc
22081 ctcagttgca attggccgag attgaaatga tcacaccgga ggagaaagta cagatcgttg
22141 aagtatttaa cgcgacatcg gctccttatc caagggacaa gaccattcat gagctgttcg
22201 cagaacaagt caagcgtaca ccagagcaga cggcgcttgt attcggcgat gtccagctaa
22261 cgtatcttga attggaagac aaggcgagcc gactggccca aacactgcgt cgtttgggaa
22321 cgttgaggga gcagcctgtg gccgtgatgg gcggacgaag catcgagatg gtcattggta
22381 tgctcgcggt gcttcaagcg ggtggagcct atgtgccgat tgatcctgat tacccggaag
22441 atcgggttcg ttatatgctt aatgattccg acgccaagct attattggtg caaaagggcg
22501 agcttataag tgtcgactac ggtataccga ttgtcgatct tagcagtgaa gaggcctatg
22561 cagctgagcc tgcccagccg gagactgctc agggatcgca ggggcttgct tatgtcatct
22621 atacatcggg tacgacgggt agaccgaagg gcgttatggt tgaacaccgg aacgtggtcc
22681 gtctagtcaa agagaccaac tatgtggagc tgaatgaatc cacacgaatt ttgcaaacag
22741 gagccgtagc ctttgatgct tctacattcg aaatatgggg agcgttgctc aacggtgggc
22801 agctctattt cgtggagaac gacgatatcc tgattgccga taggctcaaa gcggccatta
22861 ccaagtacgg gattacaaca atgtggctca cttcaccgct tttcaatcag ctttccctgc
22921 aggatgagta cctgttcaga gggctaaaag cattgctagt cggcggtgac gtactgtctc
22981 tatctcatat gaatcgtgta atcgaggcta atcctgatct tgtccctatc aattgctatg
23041 gtccgacaga gaatacgacc ttctccacca cttacaagat tccaggtcgt gccgaagggg
23101 gtgtgccgat tggtcgccca attcgtaatt cgaccgctta tgtggtcaat ggatcgctgc
23161 aattacagcc tattggtgct tggggtgaac tcattgtcgg cggtgaaggt gtagcgcgcg
23221 gatatctcaa tcgtcctgat ctcacagcag agaaatttgt tcctagtcct gtgaaggacg
23281 gagaaccctg ctacagaact ggggatttgg tacgctggct tccagatggg aatttggagt
23341 ttaaaggaag aattgatgag caggtcaaaa tacgtggtta ccgcatcgaa ctccctgaaa
23401 tcgaggccca actggccaag gtggagtcag taatcgacgc cgtagtggtc gttcgcgcgg
23461 atgagcttgg cgagaagcag ctttgcgctt attatgtggc ggatcgtacg ctcacggcag
23521 gcgaagtacg tctttcccta tcgcaggtac ttccaggcta tatgattcca tcctacttta
23581 tccagatgga tcgtatgcca ttaacgtcaa acggaaaagt ggaccgcagg tctctgccgg
23641 ctcctcaagt aggcgcgcat acaggacgga agtatacagc tcctcgtaca ccggctgagg
23701 aagctttggc atctgtctgg caaggggtgc tgggtgccaa acaggtgggt atccatgaca
23761 atttctttga attgggtgga gactccataa aagctattca ggtgtcgtca cggttactgc
23821 aggctggcta tcggttagag atgaagcagc tgttcaaatc gccaaccatt gccgagctag
23881 gcgcggaaat acaaacggct gtgcatatgg ctgaacaggg agttgtgcgt ggaacgactc
23941 gcttgactcc agtccaacag tggttctttg gacggaagca ggcagaacct catcacttca
24001 atcaagcggt tatgctgtat cgtgaacagg gatttgaaga aaaggccttg catcaggtgc
24061 taagaaaact cgctgagcat catgacgccc ttcgcatggt tttccgtcag acagagcatg
24121 gctacgaagc ttggaatcgt gatcttgatg aaggagagct ttatagcctg ttcaccgctg
24181 atttacggaa tgaatccgat ccgactgcag ccattacatc gctgtcggat gacattcagc
24241 gcagtatcaa tctggcagaa ggcccgctgc tgaagttagg acttttccat tgtcaggatg
24301 gagaccacct tctaatcgtg atccaccatt tggtggtgga cggagtatcc tggcggattt
```

24361 tgttcgagga tattgcagca ggctatgagc aggtgattca aggacaagcg ctgacattcc
24421 cgcagaagac ggattccttc cgtgactggg gagacgccct tgctcgttat tcggaaggtc
24481 ctgatatgga gactcatcgg gcgtattgga gagagctgga ggatcagcca ctcgaacagt
24541 tgccgaagga tgaggctgtg gaaagccttc ttttacagga tagcaaagta ataacagcac
24601 aatggactct agaagaaacc gaccaattgt tgagaaaagc ccatcgtgct tatcaaacag
24661 agacgaatga tctgctatta actgctctgg gcatggcggt atccaaatgg tctggcatcg
24721 gaaaggttgc tgtgaatctg gaaggacacg gtcgtgagcc gattataccg aatatcgaca
24781 tcacccgtac cgtaggctgg tttacaagtc aatttccggt gattttagac ttgggggatg
24841 acccggaagt ggcctccttg atcaagtctg tgaaagaggg gctacgccgg attccgaaca
24901 aaggtattgg ctacgggttg ctcaaaacga tggcaagtca gttggatgaa gacagcttca
24961 gcttacagcc tgagatttct tttaactatc tggggcaatt tgatcaggat ttgcaaggaa
25021 gctcgttgca gatttctcct tatccgaccg gaagcgccca aagcttgttg gaggaaccgg
25081 cctatacgct agacatcaat ggcatggtga cggacggagc cttgactctg acgattacct
25141 ataacggaaa acagtataag tcatctacga tggaacagct cgctggatat attgaagaaa
25201 gcctgcggga gcttctccag cattgcgtaa accaagaaaa aaccatattg acaccaagcg
25261 acgtgcttgc gaagggtcta agcattgccg atctggagga gctttctaag cagacgagtc
25321 acattggcga tattgagaat gtatatagtc tgacaccgat gcagaagggc atgctgttcc
25381 atgatatgtt tgagccgcat acaggtgctt attttgagca ggctgccttt gactttaagg
25441 gtagctttga tccgaccgcc ttcggacaca gtctggatgc agtggtggag cgtcatgcta
25501 tcctgcgcac gaacttttac agcggatggg gcagcgagcc tttgcaggtt gtatttcggc
25561 acagaggcgc taaattggtg tacgaagacc tgcgggagat gaatgcatcg cagcgcgaag
25621 cttatctgaa gacatttggt gctaaggaca aagcactggg cttcaaccta gctgaagacg
25681 agcttctccg tgtatcaatt ctacaaacag atgaagagag cttccgtctc ttatggagct
25741 ttcaccacat cgtcatggat gggtggtgtg ttccgttaat tacgcaggag gtatttgaac
25801 actattttgc cctcctggaa ggaagagagc cgcagttggc agaggttcat ccgtacagtc
25861 gatatatcga atggctggaa cagcaggatg aagcagctgc gtccaactat tggagccgat
25921 atctggcggg ttacgagcag cagacgcttt tacctcaagt cggtggagca agtaagggag
25981 aaggctatgt agcagaaaag ctgaattatc ctctcagcag ggaattgact gagcgccttg
26041 aaaaggtggc cagggatgcc catgtcacga tgaatatatt gctgcagtcc ctctggggca
26101 ttgcgcttca acgctataac ggtagcaagg atgtcgtgta cggaagtgta gtatcaggca
26161 gaccagcaga aattccgggc attgatcgga tgatcggttt gttcatcaat acgattcccg
26221 ttcgtgtgaa gacagaggag aatctcccct tcacagttct gatgaagcag cagcaggaac
26281 aatatatggc ttctcatatg tatgacacct acccgctgtt tgagattcag gctcagacgg
26341 atcagaagca ggatctaatc tcccatatta tggtgtttga gaactatcct gtggaagagg
26401 aggtagagcg tctgggtggt ggcgaggctg cctttgagat tgaggaagcg gagcttcttg
26461 agcaaacgaa ttatgatttt aatttaattg tcctccctgg cgaagaaatg agattgctgt
26521 tccagtacaa tgcacttgtt tatgacccag tgacaattga acaaatcaag ggccatctgg
26581 ttcacctcat ggaacaaatt gtagagaacc ctgccatttc cgtggatgca ctagaattag
26641 tcacgccgca ggagagagaa cagattctga acgtatgggg aaatacaaaa ggcatttacg
26701 agcactgtaa cacgttccac gggctgttgg aggaacaggc gggacgaacg ccggatgcga
26761 ctgccatttg gttcgaggac gagagtctga cctatgccga gctcaatgca aaagccaatg

```
26821 gactggcgag aaggctccgt actcagggaa tcaagacggg agatctggtg ggactgattg
26881 ctgaacggtc gctcgaaatg atcgttggga tctacggcat tatgaaagcc gggggtgcct
26941 atgttccaat cgatccagag tatccgaaag aacgaatcag ttacatgctt gaagattccg
27001 gggcgaagct gatccttaca caggcccatc tcttggagca tcttggatgg acggaaaatg
27061 ttttgctgct ggatgaatca tcgacctatg atgccgatac ctcgaatttg gaggatactg
27121 ctggcccgga tgatctggct tacgtgatct acacttcagg tacgacgggt cagcctaagg
27181 gcgtattagt cgagcatcgg ggactaccga atctttcgga cgtatacggg acacacttcg
27241 aagttacacc gcaggatcgg atcgttcagt ttgcaagcct gtcgtttgat gcatcggttt
27301 cggaaatttt aacggcgctg agccacgggg gtgttctgtg catcccttct acacaagata
27361 ttttagatca tgccctgttc gagcagttca tgaacgataa ggggattacg gtagcgactt
27421 tgccacccgc ttacgctatc caccttgatc cagagcgttt gccaacactg cggtgcctgc
27481 taaccgctgg atcggccgca tcggtcgagt tgatcgaaga gtggaggaag catgtacgtt
27541 actctaatgg ctatggccca acggaggact ccgtatgcac caccatctgg tctgtcccgg
27601 acagtgagga agcaacggat attgtatcta ttggacgacc tattgctaac catagtgtgt
27661 acatcttgga tgaccatttt agattgcaac ctgtaggtgt agctggagag ctatgcattt
27721 cgagtatcgg gttagcacgg gggtatcata accggcctga gttaatggat gagaagtttg
27781 tggacaatcc gtttgctcca ggagagcgta tgtatcggac gggtgacctg gttcgctggt
27841 taccgaatgg aaccatcgag tacttaggtc gaatagatca ccaagttaaa atccgcggtt
27901 accgtatcga gctgggcgag gtagaagcac aaatgctcag agtgccgtcc gttcaggaag
27961 tcgtagccat ggctgtagag ggcgatgacg gctacaaaga tctagtcgct tacttcgtag
28021 ccgctcagaa acttgaggta tccgaacttc gggctgttct gtcggagatg ttacctggat
28081 atatgatccc ttcccgcttc atacaactgg aggacatgcc tctgacgtcg aacggaaaaa
28141 tcgatcggaa agcgctgaag ggcgagcgtg gatgggcagc ggcttcatct gaggctccaa
28201 ggacacctgt ggaaatccaa ttagccgaaa tctggcaaga ggtgctgggt gtagagagcg
28261 cgggagtgaa ggataatttc ttccattttg gaggtcattc actgcgtgca gccctgctag
28321 tctcacgaat tcgcaaggaa atgaatcgcg agattagtct gagagcagtg ttcgagtctc
28381 ctactattga aggattggct cgtgccattg agggctatac accgctgaat ttcgaagaaa
28441 ttcctacagc gggagcgaga gagcattatc cattgtcctc ggcccaaaaa cgactgttta
28501 ttctaagtca gctggaaggt ggagagctga gctacaatat gccggggatc cttaccgttg
28561 agggagcttt ggatcgggaa cggctagagc aggcattccg tcgtctgatt catcgtcatg
28621 gttcgctgcg tactcgcttt gtgaccgtga acggtgaacc tgtacagcag ctcctgacag
28681 atgttccgtt tactgtggaa tatgcggagt tgagcgagga agaggcagga gctacccttc
28741 agcagtttgt ccgtcctttc gatctaggtg tagctccatt gctgcgggtt ggccttattc
28801 gaattgcaca tgagcgccat ttactattgt ttgacatgca tcatattgtc tcagatgggg
28861 tttctatgaa tattctcata gaagagtttc tccgcttcta ccaagaggag gacgtatttc
28921 ctgagctaca gatccagtac acagactatg ctgtatggca gcaagagcag ctcgaaagcg
28981 agcgtcttaa ggcccaggag gcttactggc tggatgcttt ccgtggaagc ttgccagtgc
29041 tggatttgcc aggagatgaa gttcgtcctg cggtgcgaag ctttgcgggc gatcgaatcg
29101 acttccaaat tgattcttct ctgagtgctt cacttcagga gctggctacc cggacgggtt
29161 ccactctgtt catggtactg ctggcagcct atacagcgct cttgcacaag tacacaggac
29221 aggaagatgt cattgtcggt tcacctgtgg caggaagatc ccatgcgaca cttgaaggtc
```

```
29281 tcatcggtat gttcgtcggc accgtggcac ttcgtactta tccagaagga gagaagcctt
29341 tcgaggctta tctgcaggaa gtgaaggaaa cagcgctgcg ggcctatgaa aaccaggatt
29401 acccgttcga ggagctggta gaaaagctgg agcttcagcg tgatttgagc cgtaacccgc
29461 tatttgatac catgtttgtc ctgcaaaata ttgagcaggg agaacaagaa atagaaggat
29521 tgcgcttcac tccttacgat aatgtacatc cggctgccaa gttcgatctc acgctgaccg
29581 tgagtgaagc agatggggca ttgaactgca cgcttgagta tgcgactgcg atctacaagc
29641 aagagactgc tgagcggatg gcaggccact ttgtacagct tattcgggaa gccatcgcca
29701 atccgggaat gccgttgtca tcccttgata tcgtgacacc tcaggaaaaa tcaaggctga
29761 tggaagcgcc ggacgaagcc aaggcagatt atcctcgtga caagacgatc catgcgctgt
29821 tcgaggagca ggccgcacgt actccgaatg cagtggcagt cgtatgtgaa gatgcagccc
29881 tgtcctacag cgagctgaac gagcggccca atggacttgc ccgaacgctg agggaacgtg
29941 gtttgcaacc agacggtttg gctggaatca tggcggatcg ttcccttgaa atggtggtcg
30001 gaattttagc catcttgaag gcaggcgggg cctatgtccc tgtagaccct gaatatccag
30061 aggaccgtat tcgctttatg cttgaggatt cgggagccaa gctactgctg acacaagcgc
30121 atctggagca acgtgtctcc ttcgccgggg atatcgtaag tctggacaaa acggcttcct
30181 acaaggagga tgtctcaaac ctgcagcctg cagccggacc ggagcatctt gcctacgtca
30241 tctacacatc gggtacgaca ggcaagccaa agggaacact gattgagcat aaaaatgtag
30301 ttcgcttgct ctttaatgat aaaaatatgt ttgattttgg tcctcaggat acgtggacac
30361 tgttccattc attctgtttt gacttctctg tatgggaaat gtacggagca ttgctgaacg
30421 gaggaaggtt ggtcatcgtt ccatcgctta ccgcgaagag tccagatcgt ttcttgcaat
30481 tgcttaagga tcagaaggtc accgttttga accagacacc gacgtacttc tatcagttgc
30541 tacaggaaga gcttggtcat caggcggcag aactgagcct ccgtatgatt atcttcggtg
30601 gagaggcatt agccccggcc ctgctcaagg actggagaac gaagtatccg caagtgcagc
30661 tcattaacat gtacggcatt actgaaacga ccgtgcatgt aacctacaag gaaattacag
30721 agttggaaat tgaacagggc cgcagcaata tcggcaccac gattccgacg ctgcgagcat
30781 acatttgga tgaacaacgc cgtccacagc cgattggcat tccaggtgaa ctctatgtgg
30841 cgggcgtagg tctggcgcga ggttatctga accgaccgga attgacggaa gagaagtttg
30901 tcgctcatcc gtttgaagcg ggcgagcgca tgtaccgctc gggtgacttg gcacgctggt
30961 tgccggatgg cagcatggag tatttgggac ggattgacca tcaggtaaaa atccgtggtt
31021 accgtatcga gctgggcgaa gtggaagcga agctgctcca tgctccagct gtaagggagg
31081 ccgttgtgct cgcccgggag gatggaagtg gacaaaaggt gcttgtcggc tatttcactg
31141 ccgatcagat gctgacggta ggcgagttga gaaaagcctt ggctgccgaa ctgccgactt
31201 atatgattcc atcttacttt atgcaattgg aacagatgcc tttgacgccg aatggcaagc
31261 tggatcgcaa agcgcttccg gctccggaga ccaatgtgca gactggagcg gtttatgaac
31321 cgccaaggac gaaggctgag gaagctttgg cttccgtatg gcaaggtgtg ctgggagcgc
31381 agcaggtcgg catccatgat catttctttg atctgggtgg tgattccatc aaggcaatcc
31441 aagtgtcctc gagattgttc caagccggat ataaattaga gatgaaggat ctcttcaaat
31501 atccgacaat tgccgagcta agcccgtatc ttcaggcagc cggacgtaca gcggaacagg
31561 gcgaaattaa aggtgcagca gagttaatgc caattcagcg ttggttcttt gaacgccata
31621 cagcggagcc gcatcattat aatcatgccg tcatgctcta tcggaaagac ggctttgatg
31681 aggctgcact ccggttgaca atggaccaaa ttgcgatcca tcatgatgcg ctgcgcatgg
```

```
31741 ttttccggcc tacagaagct ggatacgcag cttggaatcg gggaacggac gaaggcgagc
31801 tctacacatt ggacattgcc gatatgcggc aggcggaaga ccagacagct gcggttcagg
31861 cccaagccga tgccattcag gcaagctttg acttggaagg tggcccactg ttcaagctag
31921 gcctgttcca ttgtgacgat ggcgatcatt tgttgattgt cattcatcac ctcttggttg
31981 acggcgtatc ctggcgcatc ctgtttgagg acattgcagc ggggtacgag caggcattga
32041 atggacaagc aatcgtcctt ccacaaaaga ccgattcgta tcttgtatgg tctgagcaag
32101 cgacgaagta tgcagcaggg cctgccctgg acaaggagcg tgcatactgg caactgatcg
32161 aggaggcgat tttggcccca ctgccgaagg atgaggatca gaagccgggc accattcggg
32221 atactgaatc ggttacggta acgtggtctg cgcaggaaac agacctgctg ttgcggcaag
32281 cgaaccgggc atatcatacg gagacgaacg acttgctctt gactgctctg gggcagcca
32341 ttcagcgctg gacaggcatg gagcagattt tggtcaatct cgaagggcat ggacgggaaa
32401 tgatcgtacc agacctggat attacccgta ccgtgggttg gttcacaacc cagtatccag
32461 ttctgctgaa cctgcaaggc ggacaggaag tatctgcacg aatcaagcgc atcaaggaga
32521 atttgcgaga ggttccgcac aaaggaatcg gctacggtct tctgaagtat atggcaccgg
32581 agaaaagtgt cggatttggc gtggagccgg aaatttcctt caattatctt gggcagtttg
32641 atcaggatct ggagggtaat gcccttagct tatccacaca ttcagttggt aaggcgctca
32701 gcgacctcac accacagcaa tatgctctgg atgtgaatgg catgattgct gaaggccagc
32761 tatcactgac gattacgtac agcagcaggc agtatcgtaa cgagacggtg agtcattttg
32821 ctgaattatt acagtcaagc cttagtgagg ttatccgtca ttgtgtggct caggagcgtt
32881 cacagcttac accaagtgat gttctgttcc aagggttaac attggagcag cttgatcgac
32941 tcacggcgca gacggctcat atcggagaga ttgaggatgt gtacaagctg acgccaatgc
33001 agaagggaat gctgtttcac agcctgctgg agccggactc ctcttcatac ttcgaacagg
33061 caacgtttga gctgcgtggt agcttcgatg tagatatctt cttcgagagc ttccaggctt
33121 tggtgcaacg acatgccata ctacgtaccg gattttacaa caacattgct gatgtaccgc
33181 tgcaagttgt ctttaagcaa cggttaatcc ctctgcacta cgtagatttg cgcgacgcat
33241 ctatgcagga acaagaagcc cgaatcaagg cttatattgc tgaagatatg gttaaggggt
33301 tcagcttgtc agaagatccg ctgatgcggg tgaccgtctt gcagaaggat caaagctgtc
33361 ttgtattgtg gagcttccac catatcgtca tggatggctg gtgtatcccg atcattacac
33421 aggagctgtt cgattattat tctgccaaga aacagcaagt acagcctgtc ctgccgccag
33481 ctcagcccta tagccgttat atcgagtggc ttgatgcaca ggatgatcaa gaagcttcaa
33541 cgtattggag ccaatatctc gaagattatg acgggaatac tgtattgccg gaaggtaaaa
33601 cgaaatctca agccaaagag gcgggctatg ttctgaaaga gcatgttctc catctgggtg
33661 catctttgac aggtaaaatg gatgttgtcg cgaagcgcaa tcacgtgacc gtcaacacac
33721 tcatgcagac agcctgggga ctgattcttc aacgttataa tgccagctcg gatgtcgttt
33781 tcggcggcgt tgtgtcgggt agacccgctg agattgcagg gatcgaaaat atggtgggtc
33841 tgttcattaa tacagtacct atccgtgtac agtcatccaa agacgaagcc tttgtcgagg
33901 tgatgaaacg tacacaggca cagtcattgg ctggtcgtgc ctacgacacc tatccattgt
33961 atgagattca ggggaaaaca acccaaaagc aggacctgat ttctcatatt atgatctttg
34021 aaaattaccc gctcgacgag caggtggagc aatcgggtaa tcaaacggag gacaatctcg
34081 aagtcgccaa cttcaccatg tttgaacaaa ccaactatga ctttaacctg gttgtgattc
34141 caggcgaaga catcaaggtc tgcattcgct ataatgcttc ggtttacgag caagaaagca
```

```
34201 ttgcacgtat cggaggacac ttgttgcaga tgctcgatca ggtggctgct cgtccgcagg
34261 cgacgataca ggaactggag attgtaacat ccgaggaacg gataaacctg cttgactggg
34321 gcggcaaggc ccatacctat ccaagtgatc agggtctgca taccttgttt gaagaacagg
34381 ttgtccgtac gccggataag attgcggctg taaatgggga cactcagatc acgtatcggg
34441 agctgaacga gcaggcgaac agactagctt ccaccttgat agaccaagga ctacagagtg
34501 aacaagtggt aggtctgttg gcagatcggt ctgtagagct gcttgtcgcc atcatgggtg
34561 tactcaaagc gggtggagcc tatgtaccta ttgatcctga atatccgcag gagcggattc
34621 agtatatttt gaaggattct ggcgctgaaa ttctgctcac acagagccac ctaacggagt
34681 tggcctcttt tgagggaacg gttatggaat tggattcccc gcacatctac ggaaccgagg
34741 tggataatcc caatattcct gttggaggaa acgatctggt gtacttaatc tatacctcgg
34801 gtacaaccgg aaatccgaag ggaaccatga ttaaccacaa agggatcgtg aactacatct
34861 ggtgggccaa taaggtctat tgtgctggga aaccaacgga tttcccgttg tattcatcca
34921 tttcgtttga cttgacgatg acatcaatgt ttactccgtt aataaacgga ggaatagtac
34981 ggatttatga tggtatagat aaggcagagg tcgttcagca tattttgcgc gaaaatgcgg
35041 tcgatattct caagctgacg ccgacgcatc tcagtctcat aaaagacatg accattccag
35101 cagaaagtcg cattcaacag ctcattgtgg gtggagagaa tttgaccaca catttgtcca
35161 aaacgattac cgatctcttt ggtggcaaca tcaaaatcta caatgaatat ggaccaaccg
35221 aaacggtcgt cggctgcatg attcacctgt acaatcctgc gaaggatacg cgtgaatctg
35281 taccgattgg gttgccagca gacaatatat acatccatat tctggatgaa cagcttcgtc
35341 tcgtaccgtt aggcgtggag ggtgaaatgt acatcgccgg ggacggggta gcccgtggat
35401 atctgaaccg tcctgagctt accgcagata aattcattag aaatccgttc gcttcggaag
35461 gaaatatgta tcgcactggg gatttggctc gtcgccttcc taatggagac attgagtaca
35521 ttggacgcat tgaccaccaa gttaaaatac ggggctatcg tattgagctt ggtgagattg
35581 aggccaagct gctggacatt ccacttgtcg aggaagctct cgttgttgcg tgggcagatg
35641 ctcatgggca gaaatcgctg tgtgcttact tcgtagctaa tcgcgaaatg tctgtcagcg
35701 agctgagaga cgaactgtct gccggactgc ctgcatatat gattccgtct tacttcgtcc
35761 aactggacgt gatgcctctg acaccgaatg gtaagctgga tcgcaaggca ctgcctgaac
35821 cgaactcggg tataaaggcg ggagcagact ttaccgctcc gcgcacggat gtggagaaca
35881 ttttggcttc aatctggcag ggtgtactcg gcgtgccgct tgtcggcatt catgataatt
35941 tctttgagct tggaggtgac tcgatcaaat ccattcaagt atcttcaagg cttctccaag
36001 caggctataa gattgaaatg aaggatttgt tcggttatcc gacaattgca gagctggcac
36061 agcgcgttag tgtggtcagc cgaattgcgg accaaagcga ggtacacgga gcggtaagac
36121 tgggacctgc tcagcacaga ttcttcgatg aacagtcaat ggatctgcat cactttaatc
36181 agtcggtaat gttgtaccga cgtgatggct tcaataccga tgcgctcgcc gaggttgttc
36241 ggaaaattgc agagcatcat gatgctttac gactggtgtt ccgccaagga gagcagggat
36301 tggaggcctg gaaccggagc atgggtgagg gtgagctcta tagcctccag atccacgacc
36361 tgcgggatga gacagacccg gcttcagcaa tagaagcagg tgcggaagcc attcagcgca
36421 gcatctctct ggaggatgga cctctcttta gactgggtct gttccgctgt gcggaaggcg
36481 aacatctgtt gatcgttatt catcatctgg ctgtggatgg cgtatcctgg cgtattctct
36541 ttgaggacct gcaggatggc tacgagcagg cagcacgtgg agaagcggtc aagcttccac
36601 agaagacgga ttcgtaccgt gcatgggttg agggaatcac acaatttgcg aacagtctgg
```

```
36661 cggctgaaca agaacgcggc tattgggcag aggtagaggg agatggcttt gtccctcttc
36721 ccaaagacaa ggtagacagc gctcttctca tcaaagacag tgaggctgtc acggtgagat
36781 ggtcaccaga agagacagag cagttcctga aagaagcgaa ccgcacttac aatacggagg
36841 tcaacgatct gctcctgacg gctctgggta tggctgttca cgagtggacg ggaatcgaac
36901 gtgtaggcat ccttctggag ggacatggac gggagcctgt tgtgccggaa ctggatatta
36961 ctcgcacaat aggctggttt acaagtcaat accctgtcgc ccttgagatg ggaggggaat
37021 tggagatcgg cgccagaatc aagcacgtca aggaaggctt gcgtcgtatc ccgaacaaag
37081 gtgtcggata tggtattttg aaatatttaa gcgacggttc cgacgtctcc tccttctcgg
37141 ctgaacctga gattaccttc aactacttgg gacagttcga ccaggatctt gcaggaggga
37201 tgatggaagt atcgccttac tcagtaggac ctgaggtcag tgagcagatg gtgcagcatc
37261 aggcattgaa tattaatgga ctgattgccg aaggacagct tcaactttcg gtcagctata
37321 accgtcatca gctcgacggg gagtccgtga ctaagtttgt tggcattctg aagaaccgtc
37381 tcagcgaagt cattggacat tgcgtaagta aggaaagaac agaacttaca ccaagcgatg
37441 tactcctcaa agatatcagc ttggaaaaaa tcgaggagat agaagagcag acacggcata
37501 tcggcagtat tgaaaatatg tataaactga cgccgatgca aaaaggaatg ttgttccaca
37561 gcttgctgga gcctcattcg gaagtctact ttgagcaggc caaatttgaa attcagggag
37621 cattctatcc tgaggatttc aaacgcagct taaaatatct gatgaaacgg catgccatat
37681 tgagaaccaa tttccacgcc gggtggggcg atttccctat tcagattgtg ttcaaagaaa
37741 gagcgtgtga cttcgcatac gaggatctgc acgagctgga agccgatgaa attcaagcgc
37801 gtcttgcagc ttatactgct caggacaaag caagaggctt taatcttgct gaagaagcat
37861 tgctgcgtgt tgctattcta cgtacagcag aagaggctta ccatctgctg tggagctctc
37921 atcacatcat tttggatggc tggtgtatgc cgcttgtgct ccaggaagtg tttgagacgt
37981 atggggttct gcgtgagcaa agggagcctg agcttcctgc agctgtatcg tacagccagt
38041 atattcaatg gttggagaag caaggcgagg aagaggcatc ctcttactgg agagggtacc
38101 tggaaggcta cgagcagcag acgaagctgc cacaagccat cacacagcca tcagcaaaag
38161 cagaagccta tgtgtcggag aagctagtat tcacgttgga tgcggaattg accgatcgcc
38221 tggaacaggt ggccaaacag catcaggtga cgatgaatac attgatgcaa gcagcctggg
38281 gaatcgtgtt gcagcgctac aatagaagcc aggatatcgt cttcggaagt gtagtatcgg
38341 ggagacctgc cgagattccg ggtatcgaaa gtatgatcgg tctctttatc aatacagttc
38401 cggttcgggt tcaggccgag ggaagtgatt cgttctccca tgtgatgaaa agacagcagg
38461 aattatattt ggcaggacat gcttatgatt cctatccgct ctatgagatt caagcacaga
38521 gcgaacaaaa gcaagatttg atttctcata ttatggtgtt cgaaaattac ccggtagaag
38581 agcatctgga agagaaaatt gccagtgaag aggctgaata cagaattacg gatgttcaga
38641 tgtttgaaca gacgaattat gattttaacc tcattgtgct gccgggccgt aatctggagt
38701 tcttgtaccg ttacaatgcc cgcgtctatg atcgggagag cgtggaacgc attcaaggac
38761 acttgacgag aattctgaca agcgttgctg ttcaacctac catccgtatt gatgagctgg
38821 agttgatcac gccagaagag aaatcgcaga ttatagaggt gtggggcgat acagcagctc
38881 cttatccgcg tgagcagacc cttcacggta tatttgagga aaaagcagca ctcacaccgg
38941 atcgtacagc acttatttac ggtgaaacgg agcttaccta tggagaactt catcagcagg
39001 cgaaccgtct tgcacgtacg ctgcgtgccc aaggggtcag accggaccaa ccagtcggca
39061 tcatggtcga gcgctcgctt gagatgatca ttggtattca tgccatttta aaagctggcg
```

```
39121 gggcctatgt accgattgat ccggagttcc cagaagatcg tattcgccac atgctggagg
39181 attcgggagc gaagcttctg ctgacgaaga accatctcaa agatcgtttt ccgttcactg
39241 gcacgatcct ggcacttgat gatccgcaga tgtatcatgc ggatagctcg aatctggagc
39301 caattgcggg gccggagcat ctggcgtata tcatttacac gtcaggttca accggcaagc
39361 cgaaaggtgt aatgattgaa catcgctctg ccgtccatac gctgagtcag ttggaagctg
39421 aatatccgat gttggcaggc gaccgtttcc tgctcaaaac gacattcacc tttgacttct
39481 ccgtgccgga gctgttctgc tggttctttg gacaggggac tctcgtgatc ctgccacaag
39541 gcgtggacaa agacccgatg gcactgctag aggccgtgga tacgaaccgt atcacgcatc
39601 tcaatttggt gccgtcgatg ctcagtgtgc tcgttcaata cttgaaagaa agcggcaccc
39661 aaggattcct tactctgaaa tacctgtttg cctgcggcga gacgctgcct gccaaacttg
39721 tggaagagta ttataaagta tctccttacg cagtactgga aaacatctac ggtcctacgg
39781 aagcagccgt atatgcgact cggtatacaa cgagccttga gactgcggcc ctaacgcatg
39841 tgccgatcgg caaaccgtac gctaacgtcc aagtatggat gatggacagc gcttctcagg
39901 tatcacctgt gggggtaccg ggagaactct gcattgcggg cgaaggggta gcgcgggggt
39961 atttcaacca gccggacctg acggcagaga agttcattcc tcacccgtac aaaccgggag
40021 cacgaattta ccgaacgggc gatttggccc gatggctgcc agacgggaat attgagtatt
40081 tggggcggat cgatcaccag gtgaaaatcc gcggttaccg cattgagctg ggagaagtgg
40141 aagcacaaat tctgaaggtg ccatcggtgc aggaagcggt tgctctagca ctggctgatt
40201 ccaccggaag tactcagctt tgtgcatact ttgtggccga agaggggctt gcagcgggcg
40261 tactacgcga ggcactggcc agcgaactgc caagctacat gattccgact gctttcgtac
40321 agttggcaca aatgccgctg aatccgaacg gcaaattgga tcgtaaggcg ctgcctgcac
40381 cagaagcact cctgcggagc acagcggagt atatcccgcc gcgtacgcag acagaagtag
40441 aactagcgca gatttggacc gaggtgctcg gcgtgcagga aatcggagtc aaagatcatt
40501 tcttcgaact tgggggccat tccctgaaag tattgggatt gatccagaag atctcgtccg
40561 gcatgggtgt ccagcttccg ctccaactcg tgttcaatct tccgactgtt gaggaaatgg
40621 cacatgaaat ttccaagctg cgggcaaaag atgcccctga tgaagaggaa atggaaatta
40681 ttcacttccc gggtaaagga aggatcaaag tattttgctt ccctccgcgg gtaggatact
40741 ctctagggta ttatgagatg gctagagagc tggaaggaca ttgcgaggta tacggactgg
40801 aattcatcgg tgatcgtttc cagggtcaag acatgctgga tcgatacatg gatgccatcg
40861 tgggtattca ggctgaggga ccatatgtct tcttgggata ctcactcgga ggaaatctgg
40921 ccttcgaggt ggccaaagcc atggaaagcc gaggtcacca tgtcagcgac atcattatgg
40981 tggatgctat gagaaagatg tccaaggacg aatcgacgcc ggagcagctt gaagagattg
41041 tcgaggtggt actggacagc attggggacc agtacaaatc attcctagcc gatccagctg
41101 acagggagcg agtcaaagac aaaatgttga tctactccat ctaccgcaac gagcttatta
41161 acgcaggtga agttcagtcg aatatccatg ccttggttgc agaagacgat agtattggtc
41221 cggtgacctc atcagataaa ctgctatggc aacaggcgac acttggtcaa tacgaagaat
41281 acggagtcat cggtacacac gatgtgctgc ttgattccgg ttatgtcggg gagaatgcta
41341 aagtgctgag acagatactt ggcaaggtcg cagaagcatc atctaaaaac aagcccattt
41401 tgtcataaga tcaatcaacc ataataagtg ttgggcagca aagacgcccg ggttccttta
41461 tataggagaa acccgggcgt ctttttgtgt tttatgcaac catcaacata taatgccctt
41521 aatcatctgt gtttttagg aacggttggt tgaggaggac tgatagtggg agatgagctc
```

```
41581 acggattacc tgaatgacaa gctctggttg atcatggata atcatatgct cgctcttctt
41641 ggcaacaata aatttgctgt ttgtagagat gttgagcatc tcacgctgcc caattgccag
41701 ctttgttcga tctggtcatt tccgctcttg ctaataccga actgagtcaa atcttgctgg
41761 atgccccggg caatcacacg cacaggcaga ttccccagat gctgtccccg tatggcatct
41821 tcggtagtgc tcgctaggtc accctcctct gcaacagacc gaaagtattt agaagaagcg
41881 acaatattca caaaagattt ttatcttgat attcgacccg acccgttagc ataaaattgg
41941 ggacaggcga aaagcccctg acttttctag gagaatggat attaacgggg agggattgac
42001 cttaggccgt tccttggcat agatcttatt tgtccgtcgg gcgtcattct caggcctggc
42061 atccaccaga accaatcctg ccacttcgtc tctgtatgtc tgcgcgtaca gtcttgagta
42121 catccctccg agggaatggc ctaccagaat ataaggagct tggatgcctg cgtttttaa
42181 ggcgacatga agctcttgaa cgatgtttc cccggttcgt ggagtagcag cctcctcgct
42241 ccaggcatag cctgcccggt cataggatac aacagtggcg aatgatgaca gcttttcggg
42301 aatatctctc caggagaggc tgctttcccc gctccctgat tcaagcaaaa ttgttggaga
42361 accgttaccg agtttacgaa tatgaagtcg atagcccccg gcatcgacca attcccccgg
42421 aggtgggtat tgtgatttat ccttcttgga tgcaatccac tcatacacaa atccggagat
42481 cacgaaaagc aataagcaag agataataag aaacacaatt ctcgtccatg ttcgccttcg
42541 ttgcttcatt agaaaatcta ctcctctata cagctataat taatcattac ctaaagttgc
42601 gtgaagcaaa agctgctacc gtatggtttg tggctagctc cttatctttt tacccatcct
42661 ttctttgtca aatcatcaat attaaagaaa aaacagaata tgcacttgag ttgtaatgag
42721 taacccgcgt attcttgatc attgaatgga tatcacgtcg ttctggatgc ggataactaa
42781 gtccatttgg caaaagtgca gtgtacacag gtaaaattac aacaataaag atagtaaaaa
42841 ctattcatta aaaagggttg catttaaata tacagaatgg tatactctaa ttccggccaa
42901 gaaataaaat attgcgagcg agtgtttaaa aaaaagattc taaaaaaaga ttctaaaaaa
42961 aagtgcttgc aaggttggtt cggacatgat atgatataag agttgatgag gcgacaaact
43021 gaatcaacga ggaaagaatg tttgatcttt gaaaactgaa caacgagtga gtaaaacgat
43081 tttgttcgca aaatcaaaca agagatattt tatctcgtca gtttcaaaat gagcttatcg
43141 ctctttctat aaaccagctt cggttggtct ttaatggaga gtttgatcct ggctcaggac
43201 gaacgctggc ggcgtgccta atacatgcaa gtcgagcggg gttatataga agcttgcttc
43261 tatataacct agcggcggac gggtgagtaa cacgtaggca acctgcccac aagacaggga
43321 taactaccgg aaacggtagc taatacccga tacatccttt tcctgcatgg gagaaggagg
43381 aaagacggag caatctgtca cttgtggatg ggcctgcggc gcattagcta gttggtgggg
43441 taaaggccta ccaaggcgac gatgcgtagc cgacctgaga gggtgatcgg ccacactggg
43501 actgagacac ggcccagact cctacgggag gcagcagtag ggaatcttcc gcaatgggcg
43561 aaagcctgac ggagcaacgc cgcgtgagtg atgaaggttt tcggatcgta aagctctgtt
43621 gccagggaag aacgtcttgt agagtaactg ctacaagagt gacggtacct gagaagaaag
43681 ccccggctaa ctacgtgcca gcagccgcgg taatacgtag ggggcaagcg ttgtccggaa
43741 ttattgggcg taaagcgcgc gcaggcggct ctttaagtct ggtgtttaat cccgaggctc
43801 aacttcgggt cgcactggaa actggggagc ttgagtgcag aagaggagag tggaattcca
43861 cgtgtagcgg tgaaatgcgt agagatgtgg aggaacacca gtggcgaagg cgactctctg
43921 ggctgtaact gacgctgagg cgcgaaagcg tggggagcaa acaggattag ataccctggt
43981 agtccacgcc gtaaacgatg aatgctaggt gttaggggtt tcgataccct tggtgccgaa
```

```
44041 gttaacacat taagcattcc gcctggggag tacggtcgca agactgaaac tcaaaggaat
44101 tgacggggac ccgcacaagc agtggagtat gtggtttaat tcgaagcaac gcgaagaacc
44161 ttaccaggtc ttgacatccc tctgaccgtc tagagatagc ttccttcgac agaggagaca
44221 ggtggtgcat ggttgtcgtc agctcgtgtc gtgagatgtt gggttaagtc ccgcaacgag
44281 cgcaaccctt atgcttagtt gccagcaggt caagctgggc actctaagca gactgccggt
44341 gacaaaccgg aggaaggtgg ggatgacgtc aaatcatcat gccccttatg acctgggcta
44401 cacacgtact acaatggccg gtacaacggg aagcgaaatc gcgaggtgga gccaatccta
44461 gaaaagccgg tctcagttcg gattgtaggc tgcaactcgc ctacatgaag tggaattgct
44521 agtaatcgcg gatcagcatg ccgcggtgaa tacgttcccg ggtcttgtac acaccgcccg
44581 tcacaccacg agagtttaca acacccgaag tcggtggggt aacccgtaag ggagccagcc
44641 gccgaaggtg gggtagatga ttggggtgaa gtcgtaacaa ggtagccgta tcggaaggtg
44701 cggctggatc acctcctttc tatggagaat cgtttcctgc aatggaaaca ttcaaatatg
44761 aagcgtaagc ttcaaaaact actcactcgt tgctcagttt tgagagttca aactctcaaa
44821 aaaaagctgg ttatttacag ctttgcacct tgaaaactgg ataccgaaac gaaattgcgt
44881 tttagaatat tcctttaagc tgatcttgtg taaacaagtg aaataaaggt agcaggtaag
44941 gaaagatctt ttgcctttgg caaaaatcat tctttatcga acatcgacat tttctttctt
45001 tgaaagaaaa gtctaggtta agctacaaag agcacacgga ggatgcctag gcgccaggag
45061 ccgacgaagg acgtggcgaa caacgataag gcctcgggga gctgtaagca agctttgatc
45121 cggggatgtc cgaatgggga aacccggctg tcttcatcga cagtcacttt ctgctgaata
45181 cataggcaga atagaggcag accaggggaa ctgaaacatc taagtaccct gaggaagaga
45241 aaacaatagt gattccgtca gtagcggcga gcgaacgcgg attagcccaa accaaggagc
45301 ttgctccttg gggttgtggg acgtctcaca tggagttaca aaggaaccgg ttagatgaag
45361 aggtctggaa aggcccgcca gagaaggtaa aagccctgta gttcaaaact tgttccctcc
45421 gagacggatc ccgagtagtg cggggcacgt gaaaccccgt atgaatccgg caggaccatc
45481 tgccaaggct aaatactccc tggcgaccga tagtgaagca gtaccgtgag ggaaaggtga
45541 aaagcacccc ggaaggggag tgaaatagat cctgaaaccg tgtgcttaca agaagtcaga
45601 gccctattga tgggtgatgg cgtgcctttt gtagaatgaa ccggcgagtt acgttcccgt
45661 gcaaggttaa ggtgaagagc tgaagccgca gcgaaagcga gtctgaatag ggcgaatgag
45721 tacgtggacg tagacccgaa accgggtgat ctacccctgt ccagggtgaa ggtgcggtaa
45781 cacgcactgg aggcccgaac ccacgcatgt tgaaaaatgc ggggatgagg tgggggtagc
45841 ggagaaattc caatcgaacc cggagatagc tggttctccc cgaaatagct ttagggctag
45901 cctcggaaaa aagagtcgtg gaggtagagc actgattggg tgcggggccc gcaagggtta
45961 ccaagctcag tcaaactccg aatgccatgg acttagttcc gggagtcaga cagtgagtgc
46021 taagatccat tgtcgaaagg gaaacagccc agaccatcag ctaaggtccc caagtgtgtg
46081 ttaagtggga aaggatgtgg agttgcacag acaaccagga tgttggctta gaagcagcca
46141 ccattgaaag agtgcgtaat agctcactgg tcgagtgact ctgcgccgaa aatgtaacgg
46201 ggctaaacac accaccgaag ctatggcttg atgcaatgca tcaggggtag gggagcgttg
46261 aatgcgggtt gaaggtgtac cgtaaggagc gctggactgc attcaagtga gaatgccggt
46321 atgagtaacg aaaagatctg tgagaatcag atccgccgaa agcctaaggg ttcctgagga
46381 aggttcgtcc gctcagggta agtcgggacc taaggcgagg ccgataggcg tagtcgaagg
46441 acaacaggtc gaaattcctg taccaccgta atccgttatg agcaatgggg tgacgcagta
```

46501 gggtagtgac gcggactgat ggatgtccgt ctaagcagtg aggctggtgt gtaggcaaat
46561 ccgcacatcg taaggctggg ctgtgatggg gagcgaaaat tacagtagcg aaggtcatga
46621 tctcagactg ccaagaaaag cctctagcca ggagaaggtg cccgtaccgc aaaccgacac
46681 aggtaggcga gaagagaatt ctaaggcgcg cggaagaact ctcgttaagg aactcggcaa
46741 aatgaccccg taacttcggg agaaggggtg cctcggtagg gtgaatagcc cgagggggcc
46801 gcagtgaaaa ggcccaagcg actgtttagc aaaaacacag gtctgtgcga agccgcaagg
46861 cgaagtatac gggctgacgc ctgcccggtg ctggaaggtt aaggggagtg gtaagccttc
46921 gggcgaagct atgaaccgaa gccccagtaa acggcggccg taactataac ggtcctaagg
46981 tagcgaaatt ccttgtcagg taaattctga cccgcacgaa tggcgtaacg acttgggcgc
47041 tgtctcaacg agagatccgg tgaaatttta atacctgtga agatgcaggt tacccgcgac
47101 aagacggaaa gaccccatgg agctttactg cagcttgata ttgaatttgg gtacgatctg
47161 tacaggatag gtgggagccg tagaactttg agcgccagct tgaagggagg catccttggg
47221 ataccaccct gatcgtatct aggttctaac ttggtaccgt aaaccggtgc gaggacagtg
47281 tcaggtgggc agtttgactg gggcggtcgc ctcctaaaga gtaacggagg cgccccaagg
47341 ttccctcaga atggttggaa atcattcgaa gagtgcaaag gcagaaggga gcttgactgc
47401 gagacctaca agtcgagcag ggacgaaagt cgggcttagt gatccggtgg taccgcatgg
47461 aagggccatc gctcaacgga taaaagctac cctggggata acaggcttat ctcccccaag
47521 agtccacatc gacggggagg tttggcacct cgatgtcggc tcatcgcatc ctggggctga
47581 agtaggtccc aagggttggg ctgttcgacc attaaagcgg tacgcgagct gggttcagaa
47641 cgtcgtgaga cagttcggtc cctatctgtc gtgggcgtag gaaatttgag aggagctgtc
47701 cttagtacga gaggaccggg atggacgtac cgctggtgta ccagttgttc cgccaggagc
47761 accgctgggt agctatgtac ggaagggata agcgctgaaa gcatctaagc gtgaagcccc
47821 cctcaagatg agatttccca attagtaaga ccccttgaag acgacgaggt agataggttg
47881 ggggtggaag tgcagtaatg catggagctg accaatacta atcggtcgag ggcttatcct
47941 aaaataaaac gcaaatgagt ttcggatcca gttttcaggg tgtaacacct tgaaggtatg
48001 tacaagtgac